EIILM UNIVERSITY
SIKKIM

FINANCIAL ENGINEERING

www.eiilmuniversity.ac.in

Subject: FINANCIAL ENGINEERING **Credits: 4**

SYLLABUS

Overview of Financial Engineering
Introduction to financial engineering, Basics of Probability, Probability distribution, its properties and how it is used in business activities, Stochastic model, Monte Carlo techniques and objectives.

Overview on Financial Markets
Introduction, Market players and conventions, International monetary systems, Foreign exchange markets,

Derivatives and Security Valuation
Introduction to derivatives, Forward contracts, its types and disadvantages, profit and loss from future contracts, approaches to security valuation, calculation of values of risky securities, probability forecasting.

Capital Assets- Pricing Model
Introduction to capital assets pricing model and capital market line, Option pricing, Binomial option pricing formula.

Models of Asset Pricing
Arbitrage pricing theory, APT and its relation to CAPM, single factor model, Multiple factor model, Estimating factor models, A Case Study on Asset and Liability.

Cash Flow Engineering and Forward Contracts
Introduction, Cash Flow in different currencies, Forward contracts and currency forwards, Engineering of interest rate derivatives, Swap Engineering, Repo market strategy in financial Engineering.

Engineering of Instruments and Pricing
Replication methods and synthesis, Option mechanics, Options as volatility instruments, Black Scholes Model, Engineering convexity position, Option Engineering with application, Binomial option pricing models.

Case Study and Articles
Article on option pricing a simplified approach, Article on asset and Liability Management, Case study of large losses in derivatives markets.

Engineering of Fixed Income Securities
Connection between Swap, Bonds and FRA, article on use of derivatives to manage risk, Engineering of market volatility, Financial engineering- credit derivatives, Engineering of Equity Instruments.

Suggested Reading:

1. Principles of Financial Engineering by Salih N Neftci, Academic press New York

2. International Financial Management by Apte Tata Mc Graw –Hill
3. Financial Markets, Rates and Flows by Van Horne JC Prentice – Hall
4. Options, Futures, and other Derivatives by John C Hull PHI

COURSE OVERVIEW

Financial engineering is the application of engineering methods to financial problems. The rapidity with which corporate finance, bank finance and investment finance have changed in recent years to a greater extent has ultimately led to evolution of financial engineering. One important area of study is the design, analysis, and construction of financial contracts to meet the needs of enterprises. This field is experiencing an increased demand for professionals, especially those who are trained in both the underlying mathematics/computer technologies and finance.

Tools in probability, statistics, and optimization allow financial engineers to meet businesses' needs by measuring and managing their financial risks and by designing and analyzing sophisticated financial contracts. Markets' increasing complexity fuels demand for professionals who possess understanding of the financial problems they pose, the mathematical tools to solve these problems, and the computer skills to implement these solutions. The program offers students the training they need to succeed in a career as a financial engineer. Such careers traditionally include, among others:

- derivatives research or marketing at an investment bank and its engineering
- derivatives strategy at a corporation or investment fund
- portfolio management at an investment fund or insurer
- hedging or proprietary trading at an investment bank
- risk management

The program is also excellent preparation for any job applying operations research methodologies to financial operations.

In the light of these recent developments towards complexity in the area of financial markets and activities university has introduced this paper in the course curriculum.

FINANCIAL ENGINEERING

CONTENT

FINANCIAL ENGINEERING

CONTENT

INTRODUCTION TO FINANCIAL ENGINEERING

Objectives

Upon completion of this lesson you will be able to

- Explain the nature and scope of financial engineering
- Identify the objectives of financial engineering
- Discuss the importance and limitations of financial engineering

Good Morning/ afternoon / evening to all.

This is the first session with you in this subject. With good gesture we will start this subject and I really expect from your side a lot of contribution otherwise you may not be able to understand this subject thoroughly.

I would like to convey you – this subject is highly mathematical. You will apply all your mathematical and quantitative techniques knowledge in the financial theories, economic models and financial decisions taking. All these concepts, you have to recall very nicely before you proceed for this subject. Anyway its fine we all are in good spirit to proceed. Okay?

In a very simple term financial engineering is the process by which a portfolio is designed and maintained in such a manner as to achieve specified goals. Financial institutions use financial engineering to create complex derivative instruments.

Anyway this lesson introduces some simple financial engineering strategies. We consider two examples that require finding financial engineering solutions to a daily problem. In each case, solving the problem under consideration requires creating appropriate *synthetics.* In doing so legal, institutional and regulatory issues need to be considered.

The nature of the examples themselves is secondary here. Our main purpose is to bring to the forefront the *way* of solving problems using financial securities and their derivatives. The lesson does not go into the details of the terminology or of the tools that are used. In fact, some of you may not even be able to follow the discussion fully. There is no harm in this since these will be explained in later lessons.

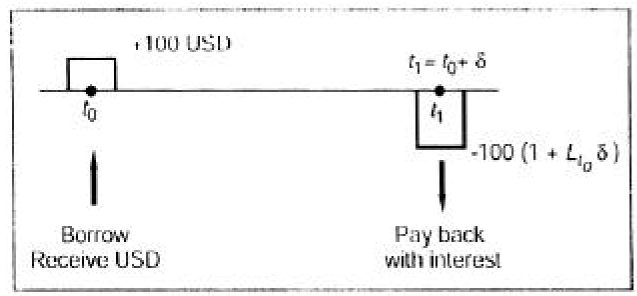
Let's Say about a Money Market Problem

Consider a Japanese bank in search of a 3-month money market loan. The bank would like to borrow U.S. dollars (USD) in Euromarkets and then on-lend them to its customers. This interbank loan will lead to cash flows as shown in Figure I-I. From the borrower's angle USD 100 is received at time to and then it is paid back with interest 3 months later at time $t_0 + \delta$. The interest rate is denoted by the symbol L_{to}¼ and is determined at time t_0. The tenor of the loan is 3 months. Therefore

$$\delta = \frac{1}{4}$$

and the interest paid becomes L_{to} ¼ . The possibility of default is assumed away. Otherwise at time $t_0 +$ there would be a conditional cash flow depending on whether or not there is default.

The money market loan displayed here is a fairly liquid instrument. In fact, banks purchase such "funds" in the wholesale interbank markets and then on-lend them to their customers at a slightly higher rate of interest.



Now You See the Problem

Suppose the above mentioned Japanese bank finds out that this loan is not available due to the lack of appropriate credit lines. The counterparties are unwilling to extend the USD funds. The question then is: Are there other ways in which such dollar funding can be secured?

The answer is yes. In fact, the bank can use foreign currency markets judiciously to construct exactly the same cash flow diagram as in Figure 1-1 and thus create a synthetic money market loan. This may seem an innocuous statement, but note that using currency markets and their derivatives will involve a completely different set of financial contracts, players, and institutional setup than the money markets. Yet, the result will be cash flows identical to those in Figure 1-1.

We can come to the solution as:

To see how a synthetic loan can be created, consider the following series of operations:

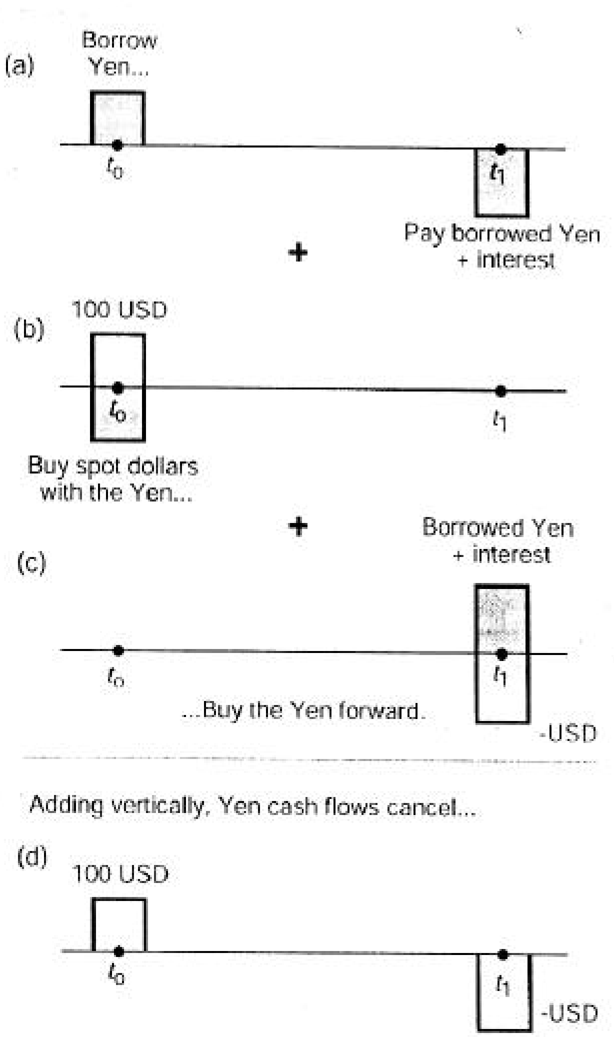
1. The Japanese bank first borrows local funds in yen in the Japanese money markets. This is shown in the Figure. The bank receives yen at time t_0 and will pay yen interest rate L^y_{to}.

2. Next, the bank sells these yen in the spot market at the current exchange rate e_{to} to secure USD 100. This spot operation is shown in the coming Figure.

3. Finally, the bank must eliminate the currency mismatch introduced by these operations: In order to do this, the Japanese bank buys $100(1 + L_{to}\delta)f_{to}$ yen at the known forward exchange rate fto' in the forward currency markets. This is the cash flow shown in the third Figure that follows. Here, there is

no exchange of funds at time to. Instead, forward dollars will be exchanged for forward yen at $t_0 + \delta$.

Now comes the critical point. In Figure two, add vertically all the cash flows generated by these operations. The yen cash flows will cancel out at time to because they are of equal size and different sign. The time $t_0 + \delta$ yen cash flows will also cancel out because that is how the size of the forward contract is selected. The bank purchases just enough forward yen to pay back the local yen loan and the associated interest. The cash flows that are left are shown in fourth Figure and these are exactly the same cash flows as in Figure one. Thus, the three operations have created a synthetic USD loan.



Here I would like to share with you some important Implications

There are some subtle but important differences between the actual loan and the synthetic. First, note that from the point of view of euromarket banks, lending to Japanese banks involves a principal of USD 100, and this creates a *credit risk*. In case of default, the 100 dollars lent may not be repaid. Against this, some capital has to be put aside. Depending on the rate of money markets and depending on counterparty credit risks, money center banks may adjust their credit, lines toward such customers.

In contrast, in the case of the synthetic dollar loan, the international bank's exposure to the Japanese bank is in the forward currency market only. Here, there is no principal involved. If the Japanese bank defaults, the burden of default will be on the domestic banking system in Japan. There is a risk due to the forward currency operation, but it is a coul1terparty risk and is limited. Thus, the Japanese bank may end up getting the desired funds somewhat easier if a synthetic is used. .

There is a second interesting point to the issue of credit risk mentioned earlier. The original money market loan was a Euromarket instrument. Banking operations in Euromarkets are considered offshore operations, taking place essentially outside the jurisdiction of national banking authorities. The local yen loan, on the other hand, is obtained in the onshore market. It would be subject to supervision by Japanese authorities. In ease of default, there may be some help from the Japanese Central Bank, unlike a Eurodollar loan where a default may have more severe implications on the lending bank.

The third point has to do with pricing. If the actual and synthetic loans have identical cash flows, their values should also be the same excluding credit risk issues. Since, if there is a value discrepancy the markets will simultaneously sell the expensive one, and buy the cheaper one, realizing a windfall gain. This means that synthetics can also be used in pricing the original instrument. [2]

Fourth, note that the money market loan and the synthetic can in fact be each other's-hedge. Finally, in spite of the identical nature of the involved cash flows, the two ways of securing dollar funding happen in completely different markets and involve very different financial contracts. This means that legal and regulatory differences may be significant.

Again let's relate it to an Example of Taxation.

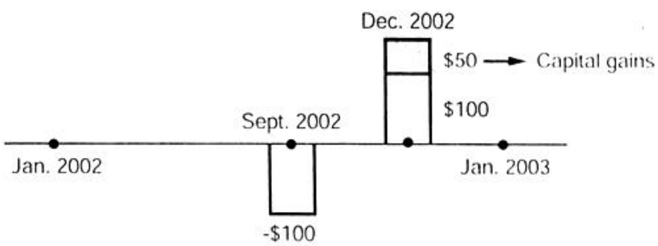
Now consider a totally different problem. We create synthetic instruments to restructure taxable gains. The legal environment surrounding taxation being a complex and ever-changing phenomenon, this example should be read only from a financial engineering perspective and not as a tax strategy. Yet the example illustrates the close connection between what a financial engineer does and the legal and regulatory issues that surround this activity.

The Problem Is:

In taxation of financial gains and losses, there is a concept known as a *wash-sale*. Suppose that during the year 2002; an investor realizes some financial gains. Normally, these gains are taxable that year. But a variety of financial strategies can possibly be used to postpone taxation to the year after: To prevent such strategies, national tax authorities have a set of rules known as washsale, and *stra4dle* rules. It is important that professionals working for national tax authorities in various countries understand these strategies well and have a good knowledge of financial engineering. Otherwise some players may rearrange their portfolios, and this may lead to significant losses in tax revenues. From our perspective, we are concerned with the methodology of constructing synthetic instruments. This example will illustrate another such construction.

Suppose that in September 2002, an investor bought an asset at a price So = \$100. In December 2002, this asset is sold at S1 = \$150. Thus, the investor has realized a capital gain of \$50. These cash flows are shown in Figure 1-3. The first cash flow is negative and is placed below the time axis because it is a payment by the investor. The subsequent sale of the asset, on the other hand, is a receipt, and hence is represented by a positive cash flow placed above the time axis. The investor may have to pay significant taxes on these capital gains. A relevant question is then: Is it possible to use a strategy that postpones the investment gain to the next tax year?

One may propose the following solution, however, is not permitted under the wash-sale rules. This investor is probably holding assets other than the S_t mentioned earlier. After all, the right way to invest is to have diversifiable portfolios. It is also reasonable to assume that if there were appreciating assets such as S_t, there were also assets that lost value during the same period. Denote the price of such an asset by Z_t. Let the purchase price be Z_0. If there were no wash-sale rules, the following strategy could be put together to postpone year 2002 taxes.



Sell the *Z-asset* on December 2002, at a price Z_1, $Z_1 < Z_0$, and. the next day, buy the same Z_t at a similar price. The sale will result in a loss equal to

$$Z_1 - Z_0 < 0 \quad (2)$$

The subsequent purchase puts this asset back into the portfolio so that the diversified portfolio can be maintained. This way, the losses in Z_t are recognized and will cancel out some or all of the capital gains earned from S_t. There may be several problems with this strategy, but one is fatal. Tax authorities would call this a wash-sale (i.e. a sale, that is being'. intentionally used to "wash" the 2002 capital gains) and would disallow the deductions.

Another Strategy

Yet investors can find a way to sell the Z-asset without having to sell it in the *usual* way. This can be done by first creating a *synthetic* Z-asset and then realizing the implicit capital losses using this synthetic, instead of the Z-asset held in the portfolio. Suppose the investor originally purchased the Z-asset at a price Z_0 = \$100 and that' asset is currently trading at Z_1 = \$50, with a paper loss of \$50. The investor would like to recognize the loss without directly selling this asset. At the same time the investor would like to retain the original position in the Z-asset in order to maintain a well-balanced portfolio. How can the loss be realized while maintaining the Z-position *and* without selling the Z_t ?

The idea is to construct a proper synthetic. Consider the following sequence of operations:

- Buy another Z-asset at price Z_1 = \$50 on November 26, 2002.
- Sell an at-the-money call on *Z* with expiration date December 30, 2002.
- Buy an at-the-money put on *Z* with the same expiration.

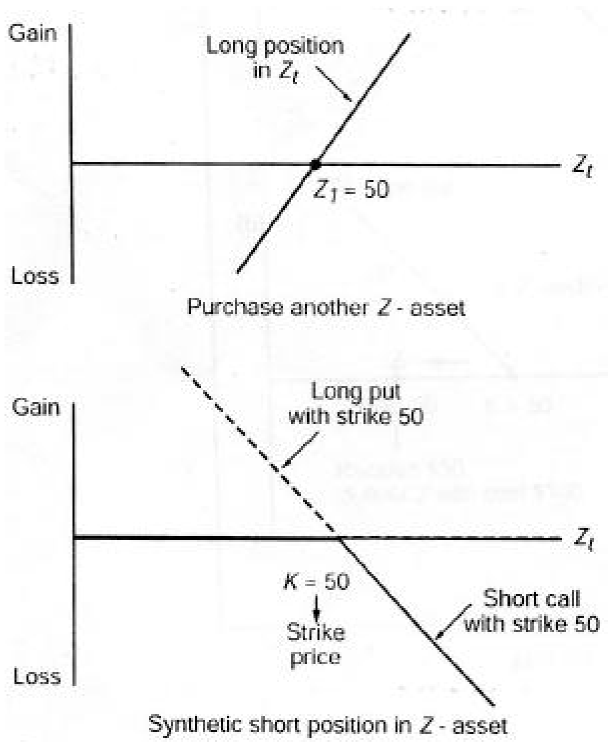
The specifics of call and put options will be discussed in later chapters. For those readers' with no background in financial instruments we can add a few words. Briefly, options are instruments that give the purchaser a *right.* In the case of the call option, it is the right to *purchase* the underlying asset (here the Z-asset) at a pre-specified price (here \$50). The put option is the opposite. It is the right to sell the asset at a pre-specified price (here \$50). When one sells options, on the other hand, the seller has the *obligation* to deliver or accept delivery of the underlying at a pre-specified price.

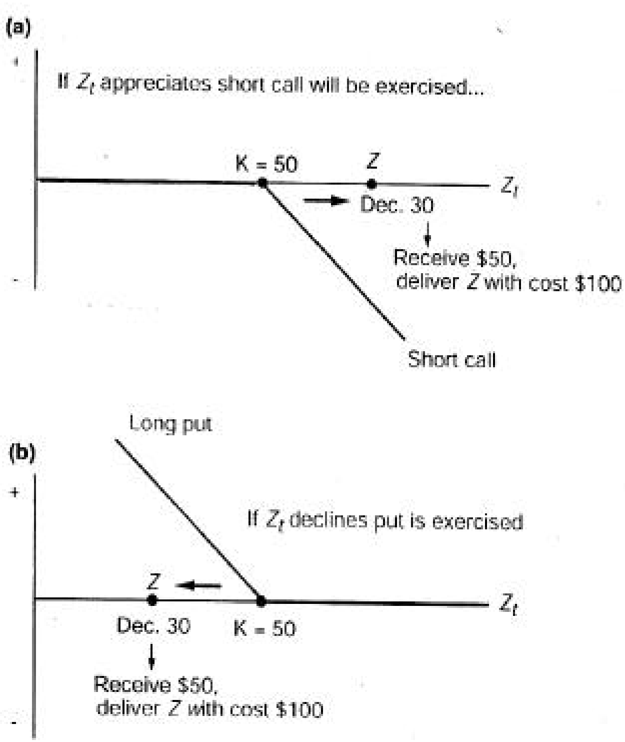
For our purposes, what is important is that short call and long put are two securities whose expiration payoff, when added will give the synthetic short position shown in Figure 1-4. By selling the call the investor has the *obligation* to deliver the Z-asset at a price of \$50 if the call holder demands it. The put, on the other hand, gives the investor the *right* to sell the 2 -asset at \$50 if he or she chooses to do so.

The important point here is this: When the short call and the long put positions shown in Figure 1 -4 are added together the result will be equivalent to a *short* position on stock Z_t. In fact, the investor has created a *synthetic* short position using options.

Now consider what happens as time passes. If Z_t appreciates by December 30, the call will be exercised. This is shown in Figure 1 -5a. The call position will lose money, since the investor has to deliver, at a *loss,* the original & Z stock that cost \$100. If, on the other hand, the Z_t decreases, then the put position will enable the investor to sell the original *Z* -stock at \$50. This time the call will expire worthless.3 This situation is shown in Figure l-5b. Again, there will be a loss of \$50. Thus, no matter what happens to the price Z_t, either the investor will deliver the *original* Z-asset purchased at a price \$100, or the put will be exercised and the investor will sell the original *Z-asset* at \$50. Thus, one way or another, the investor is using the original asset purchased at \$ 100 to close an option position at a loss. This means he or she will lose \$50 while keeping the same & -position, since the *second* &, purchased at \$50, will still be in the portfolio.

The timing issue is important here. For example according to U.S. tax legislation, wash-sale rules will apply if the investor has acquired or sold a *substantially identical* property within a 3 I -day period. According to the strategy outlined here the second *Z* is purchased on November 26, while the options expire on December 30. Thus, there are more than 31 days between the two operations.





Implication

There are at least three interesting points to our discussion. First, the strategy offered to the investor was risk-free and had zero cost aside from commissions and fees. Whatever happens to the new long position in the Z-asset, it will be canceled by the synthetic short position. This situation is shown in the lower half of Figure 1-4. As this graph shows, the proposed solution is risk less. The second point is that, once again, we have created a synthetic, and then used it in providing a solution to the tax problem. Finally, the example displays the crucial role legal and regulatory frameworks can play in devising financial strategies. Although this book does not deal with these issues, it is important to understand the crucial role they play at almost every level of financial engineering.

Some Caveats for what is to follow for your proper understanding

A newcomer to financial engineering usually follows instincts that are harmful for good understanding of the basic methodologies in the field. Hence, before we start, we need to layout some basic rules of the game that should be remembered throughout the book.

1. This book is written from a market practitioner's point of view. Investors, pension funds, insurance companies, and governments are clients, and for us they are always on the other side of the deal. In other words, we look at financial engineering from a trader's, broker's and dealer's angle. The approach is from the manufacturer's perspective rather than the viewpoint of the user of the service. This premise is crucial in understanding some of the logic discussed in later chapters.

2. We adopt the convention that there are two prices for every instrument unless stated otherwise. The agents involved in the deals often quote two-way prices. In economic theory, economic agents face the law of one price. The same good or asset cannot have two prices. If it did, we would then buy at the cheaper price and sell at the higher price.

   Yet in financial markets, there are two prices: one-price at which the financial market participant is willing to buy something from you, and another one at which the financial market participant is willing to sell the same thing to you. Clearly, the two cannot be the same. An automobile dealer will buy a used car at a low price in order to sell it at a higher price. That is how the dealer makes money. The same is true for a financial market practitioner. A swap dealer will be willing 10 buy swaps at a low price in order to sell them' at a higher price later. In the meantime, the instrument will be kept in the inventories, just like the used car sold to a car dealer..

3. A financial market participant is not an investor and never has "money." He or she has to secure funding for any purchase and has to place the cash generated by any sale. In. this book, almost no financial market operation begins with a pile of cash. The only "cash" is in the investor's hands, which in this book is on the other side of the transaction.

   It is for this reason that market practitioners prefer to work with instruments that have zero-value at the time of initiation. Such instruments would not require funding and are more practical to uses. They also are likely to have more liquidity.

4. The role played by regulators professional organizations, and the legal profession is much more important for a market professional is much more important for an investor. Although it is far beyond the scope of this book, many financial engineering strategies have been devised for the sole purpose of dealing with them.

Remembering these premises will greatly facilitate the understanding of financial engineering.

BASICS OF PROBABILITY

Objectives

- Explain you the basic definitions terminologies of probability and statistics
- Help you of how its ideas are important for financial decisions

Meanings of Different Terminologies

Hello!

In the previous class we have highlighted about the financial engineering and how financial engineering requires the knowledge of mathematics and statistics. Here let's come to the basics of probability.

As we know, statistics is the science of decision making in the face of uncertainty. The statistician is generally interested in drawing conclusions or inferences from experiments which involve uncertainties. For example, with personal observation and wise thinking, we generally make statements like, "Probably it will rain today"; The chances of teams A and B winning a certain match are equal"; "Shyam will probably gain this time from his stock market investment"; Seventy percent post-graduate students are likely to be married within 2 years" or "The chance of an examinee guessing a correct answer to a certain multiple choice question is 20 percent". In all such statements of the past experience or from an understanding of the structure of the experiment we have some degree of confidence in the validity of our statements. For making such statements related to a conclusion or inferences which have validity, an understanding of probability theory is essential for all of us. The theory of probability provides a means of getting an idea of the likelihood of occurrence of different events resulting from a random experiment in terms of quantitative measures ranging between zero and one. And you see, the probability will be zero for an impossible event and one for an event which is certain to occur. The other degrees of uncertainties or the likelihood of occurrence of events are indicated by probabilities ranging between zero and one. In other words, probability is a concept which numerically measures the degrees of uncertainty and therefore to certainty of the occurrence of events.

We can recall here of how the theory of probability has been originated. It's because, in the games of chance related to gambling with throwing a die, tossing a coin or drawing cards from a pack. Starting with games of chance, probability has become a part of our everyday life. Probability theory is now being applied in the analysis of social, economic and business problems. It has a commercial application in the field of insurance where precise knowledge of risk to life or loss is required to enumerate premium. In various activities of business, we face uncertainty and use probability theory for making management decisions. It is an essential tool in 'statistical inference' and forms the basis of 'Decision Theory'.

In fact statistics and probability are very much interrelated and it is difficult to understand statistics without the knowledge of probability.

But for understanding of the concept of probability theory, you have to grasp the following concepts and terminologies very clearly.

Terminology Used in Probability

Random Experiments

Say, if in each trial of an experiment conducted under identical conditions the outcome is not unique but may be any of the possible outcome then such an experiment is called *a* **random experiment**. For example if you toss a fair coin, then you may get a head or a tail. Another experiment may be tossing a die in which there is six possible outcome or events *i.e.* the turning up of any of the six numbers 1,2,3,4, 5 or 6. Further, the results of random experiment are called outcome or events. Thus tossing a fair coin has two possible outcome Head up or Tail up. Similarly in drawing a card from a well shuffled pack of 52 cards, there are 52 possible outcomes.

Equally Likely Events

Two or more events are said to be equally likely or equally possible if any of them cannot be expected to occur in preference to others. In other words events are called equally likely if the likelihood of the occurrence of every event is the same. For example, in tossing an unbiased coin, the head and tail has an equal chance of turning up. Similarly, in throwing an unbiased die, all the possible outcomes 1, 2,3,4,5 or 6 are equally likely.

Mutually Exclusive Events

Two events are called mutually exclusive when occurrence of one implies that the other cannot expect to occur in preference to others. In other words, events are called mutually exclusive if the occurrence of one precludes the occurrence of the others. For example, in tossing a coin either head or tail occurs, i.e., the two events, head and tail, cannot occur simultaneously. Similarly in throwing a die the occurrence of any number excludes the occurrence of the others and as such these six events too are mutually exclusive.

Favorable and Unfavorable Cases

The outcomes in an experiment which are favorable to an event in which we are interested are called favorable cases and all other outcomes are known as unfavorable cases. The sum of the favorable and unfavorable cases is equal to the exhaustive number of events in an experiment. For example, Suppose when a dice is thrown and we wish to know the probability of the event that 3 or 6 turn up. Then the two cases, *i.e.*, turning up of 3 or6, are favorable to our desired event, while turning up of 1, 2, 4 or 5 are four cases unfavorable to the event. Similarly, if a card is drawn from a well shuffled pack of 52 cards, we may be interested in the event that it is a king. Thus there are four cases (drawing of any of the 4 kings) favorable to the desired event and the remaining 48 cases are unfavorable to the desired event.

Simple Events

An event is said to be simple if it corresponds to a single possible outcome of an experiment. In simple events we consider the occurrence or non-occurrence of single events, i.e., a simple event cannot be decomposed into a combination of other events. Thus, in a single toss of a coin, the event of getting head is a simple event. In tossing two coins simultaneously, the event of getting two heads is simple. Similarly, drawing of a particular card from a pack is a simple event.

Compound or Joint Events

We can say that the joint occurrences of two or more simple events are called a compound event. Thus, compound events imply the simultaneous occurrence of two or more simple events. For example, in tossing of two fair coins simultaneously, the event of getting 'at least one head' is a compound event as it consists of joint occurrence of two simple events. viz., A: one head appears and B: two heads appear. Similarly, if "A" contains 6 white and 6 red balls and we draw 2 balls at random, then the event that 'both are white or 'one is white and one is red' are compound events.

The compound events may be further classified as

1. Independent Events.
2. Dependent Events.

Independent Events

If two or more events occur in such a way that the occurrence of one does not affect the occurrence of another, they are said to be independent events. For example if a coin is tossed twice, the results Of second throw would in no way be affected by the result of the first throw. Similarly, if a 'bag contains 5 white and 7 red balls and then two balls are drawn one by one in such a way that the first ball is replaced before the second one is drawn. In this situation the two events, 'the first ball is white' and 'second ball is red will be independent, since the composition of the balls in the bag remains unchanged before a second draw is made.

Dependent Events

If the occurrence of one event influences the occurrence of the other, then the second event is sai1 to be dependent on the first. For example, in the above example, if we do not replace the first ball drawn this will change the composition of balls in the bag while making the second draw and therefore the event of, drawing a red ball' in the second draw will depend on event (first ball is red or white) occurring in first draw Similarly, if a person draws a card from a full pack and does not replace it, the result of the draw made afterwards will be dependent on the first draw.

Definitions of Probability

We will discuss the following definitions of probability-

1. Mathematical, classical or a-priori probability.
2. Statistical, empirical or a-posteriori probability.
3. Axiomatic approach to probability.

Mathematical or Classical or 'a-Priori' Probability

If, consistent with the conditions of an experiment, there are *n* exhaustive, mutually exclusive and equally likely cases and of them m are favorable to the occurrence of an event *A*, then the probability of happening of the event *A*, denoted as *P (A)*, is

$$P(A) = \frac{m}{n} = \frac{\text{Number of favorable cases}}{\text{Number of exhaustive cases}} \quad (1)$$

And the probability that the event A does not happen will be

$$P(\overline{A}) = \frac{n-m}{n} = \frac{\text{Number of cases unfavorable to the event A}}{\text{Exhaustive number of cases}} \quad (2)$$

Clearly,

$$P(\overline{A}) = 1 - \frac{m}{n} = 1 - P(A) \quad (3)$$

or $P(A) = 1 - P(\overline{A})$

or $P(A) + P(\overline{A}) = 1$

Thus, if we know the probability of an event *A*, then the probability of its **complementary event.** *A*, is given by formula (3). From the mathematical definition of probability, it is clear that

1. The probability of an event is the ratio of the number of favorable cases to the exhaustive number of cases in a trial.
2. **Since o** ≤ m ≤ n, the probability of an event is a positive quantity, i.e., P (A) e" 0.
3. The probability of an impossible event is zero.
4. The probability of a sure event is 1 and, therefore, the probability of happening of an event ranges between 0 and 1, i.e.,

   O≤ P (A) ≤ 1.
5. The swn of the probabilities of happening and non-happening of an event is always equal to one, i.e.

   P (A) + P (A) = 1. Odds in favor and odds against

Probabilities of happening and non-happening of an event can also be expressed in terms of' odds'. The odds tl1at an event will occur are given by the ratio a/the probability that the event will occur to the probability that it will not occur. These are usually expressed in terms of positive integers having no common factor. Therefore, if the odds are a: b in favor of an event A, then

$$P(A) = \frac{a}{a+b} \text{ and } P(A) = 1 - P(\overline{A}) = 1 - \frac{a}{a+b} = \frac{b}{a+b}.$$

Thus, to say that odds are a: b in favor of an event is the same as to say that odds are b: a against the event. We will discuss it with many examples:

Example 1. If odds are 2 to 3 in favor of event A. Then find the probability of occurrence of A. Also find P(A).

Solution: Since odds in favor of event *A* are 2: 3,

$\therefore$ P(A) = $\frac{2}{2+3} = \frac{2}{5}$ and P (A) = 1- P $(\overline{A}) = \frac{3}{5}$

Example 2. The odds against an event A are 5: 8. Find the probability of occurrence of the event A. Solution: Since the odds against the event A are 5: 8, so the probability of non - occurrence of A, i.e.,

P(A) = $\frac{5}{5+8} = \frac{5}{13}$.

Therefore, P(A) = 1 – P(A) 1 – $\frac{5}{13} = \frac{8}{13}$

Let's See The Limitations of the Classical Definition Of Probability

The classical definition of probability fails to give the probability of an event in the following cases:

1. When various outcomes of a trial are not equally likely: For example, if a dice is so biased that it gives even numbers more often than odd numbers then the occurrence of numbers on the dice is not equally probable while this is a necessary condition for calculating probability with this definition. Similarly, we cannot apply this definition to find the probability of a success of a student in an examination just because the event of, success' and 'failure' are not equally probable.
2. If the exhaustive number of cases *(n)* in a trial is infinite: In this case also the classical definition fails to give the required probability. For example, in considering the probability that bulb will burn less than 1500 hours, we have no way of enumerating the total number of cases and number of favorable cases.
3. It may not be possible practically to enumerate all the possible outcomes of a certain experiment and, as such, the definition fails to give a measure of probability.
4. The classical definition of probability depends on 'equally likely' cases, which means cases 'With equal probability'. What it means is to define probability in terms of probability and the definition becomes circular in nature, which is clearly unjustified.

Statistical or Empirical or a-posteriori Definition of Probability

The classical definition of probability requires that n is finite and that all cases are equally like. These are very restrictive conditions and, as such, cannot cover all the situations. For overcoming SU (situations, the statistical or empirical definition of probability is useful. According to this definition, (m/n) is the relative frequency or frequency ratio of an event A connected with a random experiment, them the limiting value of the ratio as n increases infinitely is called the probability of the event A Symbolically,

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

For example, consider a coin tossing experiment and let *A* be the event that a throw results in a head, If the coin is tossed 10, times resulting in 6 heads and 4 tails. The relative frequency of head is thus $\frac{6}{10}$ = 0.6. However, if the experiment is carried out a very large number of times we expect that 10 the relative frequency of heads will become stable and tend towards 0.50. This indicates that through the results of an individual experiment are unpredictable; the average results of a long sequence of random experiments show a very striking regularity and are somewhat predictable.

Following illustrations will clarify the computation of probability according to Statistical or Empirical definition of probability.

Example 3. Suppose we have the following information about the distribution of marks of 1000 students in a college.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	6O-1	Total
No. of Students	40	50	150	400	230	100	30	1000

Now a student is selected at random, then let us finds out the probability that his marks are (i) under 30, (ii) above 50 and (iii) between 20 and 50.

In case (i) there are 40 + 50 + 150 = 240 students scoring under 30 out of the total number of 1000 students.

:. P (A = the event that the student selected has marks less than 30) = $\frac{240}{1000}$ =0.24

In case (ii), the number of student scoring above 50 is 100 + 30 = 130. Thus,

P (B = the event that the selected student has marks above 50) =$\frac{130}{1000}$ = 0.13

In the last case (iii), the number of students scoring between 20 and 50 are 150 + 400 + 230 = 780.

Therefore,

P (C = the event that the selected student has marks between 20 and 50 = $\frac{780}{1000}$

= 0.780.

Example 4. In the following bi-variate frequency table we are given the distribution of marks of 250 students in Economics and Statistics.

	Marks in Statistics				
Marks in Economics	20-30	30-40	40-50	50-60	Total
20-30	30	10	10	0	50
30-40	20	60	10	20	110
40-50	10	20	20	10	60
50-60	0	10	10	10	30
	60	100	50	40	250

Now if a student is selected at random from the above 250 students then let us consider the probability that (i) the selected students has his Economics marks in the group (20-30) and scores more than 40 in Statistics. (ii) The selected student's marks in Statistics are in the group (40-50) and his score in Economics is higher than 40.

In case (i), let *A* be the event that the selected student has marks ill Economics in group (20-30) and scores more than 40 in Statistics. From the table, it is clear that the number of students having such characteristics is 10 + 0 = 10. Therefore

$$P(A) = \frac{10}{250} = 0.040.$$

Similarly, let *B* be the event that the randomly selected student has marks in Statistics in group (40-50) and his score in Economics is higher than 40. From the table, the number of such students is 20 + 10 = 30. Thus

$$P(B) = \frac{30}{250} = 0.120$$

Merits and Demerits of Statistical Definition of probability

The statistical or empirical definition of probability is not suitable from mathematical point of view as it depends on the actual working and observed numerical data. However, from a practical point of view this approach is very useful in many situations. For example, according to 1991 census there are 406 million females in a total population of 843 million. The probability that a randomly chosen person is a female is = 0.482 According to classical definition the probability that a person chosen at random is a female is 0.5, whereas the probability measure 0.482 for this event obtained by statistical definition seems more reasonable.

In this way we have seen two approaches to the definition of probability. The mathematical definition does not require the results of a certain number of repeated trials of the experiment. On the other hand, statistical definition depends on the results of a certain number of repeated trials of the experiment and hence it is called a-posteriori or empirical definition. Here the resulting empirical data of the experiment is the basis of calculating probability. The statistical definition of probability also removes all the limitations of the mathematical definition. The only limitation of the statistical definition is that it is difficult to prove the existence of a limit to the relative frequency.

Next we will discuss axiomatic approach to the probability. However, for its proper understand-ing, we need to study the concept of Set and certain set operation which we will discuss in the next class.

Set Theory

Sets and Elements

Any well desired list or collection of objects is called a set. Each object in a set is called an element of the set or a member of the set. Sets are usually denoted by capital letters *A*, *B*, *X* or *Y*, whereas small letters a, *b*, *x* or yare used to indicate elements of a set. For example the set A, of possible outcomes when a die is tossed may be written as

$A = \{1, 2, 3, 4, 5, 6\}$

Similarly, the set *B* of vowels may be written as

$B = \{a, e, i, o, u\}$.

In set notations the symbol º means "is an element or 'belongs to' and º means "is not an element of ' or 'does not belong to'. For example, if *x* is an element of set *A* and y is not, then symbolically we write

x º *A* and *y* º *A*.

Similarly, for the set *A* = {2, 3, 4}, 3 º *A* but 7 º A.

Equal or Identical Sets

If two sets A and B are equal or identical, then every element of A is an element of B, and every element of B also belongs to A. symbolically, we write A = B. For example the sets

A= {2, 4, 8}; B= {2, 4, 8}; X= {2, 2, 4, 8}

are equal and we say A = B and A = B= X Here it is notable that repetition of an element in a set is meaningless. . However, if we consider a set C = {1, 2, 3, 8, 4}, then A = C and B = C.

Null or Empty Set

A set that contains no elements is called Null or Empty set. In notation, a null set is written as *f*

For example, in a die tossing, experiment, the set X, of numbers greater than 7 is a null set.

Subset and Proper Subset

If every element of a set A is also an element of B, then A is called a subset of B.

Symbolically, A Ì B denotes that A is a subset of B. Thus (i) every set is a subset of itself (ii) the null set f is a subset of all the sets. However, any subset of a set which is not the set itself is called a proper sub-set of the set. Thus, B is a proper subset of A if BÌA and B = A. For example, the set B = {1, 2, 3} is a proper subset of the set A = {1, 2, 3,4}. However, the set X= {2, 1,3,4} is a subset of A but not a proper subset, because A = X

Set Operations

There are certain operations on sets which result in the formation of new sets.

Intersection of Two Sets

The intersection of two sets *A* and *B* is the set of elements that are common to both *A* and *B*.

Symbolically, we write $A \cap B$ for the interaction of *A* and *B*. For example, let us consider the sets

A= {1,2,3,4,5,6} and B= {5,6, 7,8,9}.

Then, the intersection of two sets *A* and *B* will be

$A \cap B = \{5,6\}$.

if *A*= {3,3,4,5,2} and B= {6, 7, 1,7,6},then

$A \cap B$ = f = Null set as there is no common element in *A* and *B*.

Disjoint Sets

In case two sets have no common elements or, in other words, the intersection of two sets is a null set, the two sets are called disjoint sets. Thus, *A* and *B* are disjoint if $A \cap B = f$ For example, if *A* = {*a, e, i, o, u*} and *B* = {*m, n, p, k*}', then A Ç *B* = *f* , so *A* and *B* are disjoint sets.

Union of two sets

The union of two sets *A* and *B* is the set of elements that belongs to *A* or to *B* or to both. Symbolically, we write $A \cup B$ for the union of *A* and *B*. For instance, let us consider the sets *A* = {2, 3, 8, 5} and *B* = {8, 6, 3,4}; then the union of the sets *A* and *B* will be

$A \cup B$ = {2, 3,4,5,6, 8}

Complement of a Set

If *A* is a subset of the universal set *U*, then the complement of *A* with respect to *U* is the set of all elements of *U* that are not in *A*. The complement of *A* is denoted by *A*' . For example, consider the universal set *U*= {1, 2, 3, 4, 5, 6, 7, 8, 9} and set A = {1, 2, 3,4, 5}; then *A*' = {6, 7, 8, 9}. Similarly, *if U*= {all 52 cards in a pack} and *A* = {all red cards in the pack}; then A' = {all card that are not red, *i.e.*, all black cards}.

Fundamental Rules of Counting

In computing probabilities of complex events, it is often difficult to count the number of favourable or exhaustive cases. To facilitate the labour involved, we discuss a few fundamental rules of counting.

Rule 1

If an event can happen in anyone of m ways and when this has occurred another event can happen in anyone of n ways, then the number of ways in which both events can happen in the specified order is m x n = mn. Let us put it the other way. If an operation can be performed in m ways and if for each of these a second operation can be performed in n ways. Then the two operations can be performed together in m x n = mn ways.

Example 5. Let us count the number of exhaustive cases when a pair of dice is thrown once. In this case the first die can land in anyone of six ways. For each of these six ways the second can land in 6 ways. Therefore, the pair or dice can land in 6 x 6 = 36 ways.

Example 6. If there are three candidates for president ship and four for vice-president ship in a union election of a college. Let us find the number of ways the two offices can be filled. Since the office of the president can be filled by any of the three candidates and for each of these 3 ways, the office of the vice president can be filled in 4 ways. Therefore... the two offices can be filled in 3 x 4 = 12 ways.

Rule 2

If an event A can occur in total of m ways and if a different event B can occur in n ways, then the event A or B can occur in m + n ways provided the two events are mutually exclusive (cannot occur simultaneously).

Example 7. In a certain class a class representative is to be chosen from 3 female and 4 male candidates. Count the ways in which a class representative can be chosen.

Here a female representative can be chosen in 3 ways and a male in 4 ways. Therefore, the number of ways in which a class representative can be chosen will be 3 + 4 = 7.

Example 8. *A* bag contains 6 red, 4 white and 3 blue balls. Count the number of cases in which a drawn ball at random is either red of white. .

A red ball can be drawn in 6 cases and a white in 4 cases. Therefore, the required number of cases will be 6 +4 = 1.0.

Factorial Symbol

In the following rules we will observe that the products of consecutive integers are involved. We represent this product by a factorial symbol. For example, the product 5 x 4 x 3 x 2 x I is written as 5! and referred to as '5 factorial'. In general, for any positive integer *n*, the product *n (n* - 1) *(n* - 2)...(3) (2) (1) is represented by the symbol *n!*, which is read as 'n factorial'. By definition I! = 0! = 1.

Permutation

A permutation is an arrangement of all or part of a set of objects. Consider the three letters a, band c. The possible permutations of these three letters are abc, acb, bac, bca, cab, cba Thus we arrive at 6 different arrangement of three letters or objects. Using rule I, we could have arrived at the result without actually writing the different orders. Here, there are 3 positions to be filled from the three letters. Thus, we have 3 choices for the first position, 2 for the second, leaving only 1 for the "last position, giving a total of 3 x 2 x.l = 6 permutations. In general, the number of permutations of n distinct objects will be

$$n(n - 2)\ (n - 1)...(3)\ (2)\ (1) = n!$$

Permutations of n Objects Taken r at a Time

The number of permutations of the three letters a, band c will be 3! = 6. Let us consider now the number of permutations that are possible by taking the 3 letters 2 at a time. These permutations would be ab, ac, ba, ca, bc, cb. Applying role 1 again, we have 2 positions to fill with 3 choices for the first and 2 choices for the second, i. *e.*, a total of 3 x 2 = 6 permutations. In general, *n* distinct object taken *r* at a time can be arranged in *n (n* - 1) *(n - 2)... (n - r* + 1) ways. This product is represented by the symbol

$${}^{n}P_{r} = n(n-1)\ (n-2)\ \dots\ (n-r+1) = \frac{n!}{(n-r)!}.$$

Example 9. How many ways 5 students be lined up to get on a bus. Using (i) the total number of such permutations would be 5! = 5 X 4 X 3 X 2 X 1 =120

Example 10. How many ways can the 4 starting positions in a team be filled with 9 students who can play at any of the positions.

Here it is a problem of arranging 9 students taking 4 at a time. Using (ii) we have

$$^9P_4 = \frac{9!}{5!} = \frac{9 \times 8 \times 7 \times 6 \times 5!}{5!} = 3024$$

Example 11. Two lottery tickets are drawn from 25 for first and second prize. Count the total number of arrangements of the tickets.

The total number of such arrangements will be

$$^{25}P_2 = \frac{25!}{23!} = \frac{25 \times 24 \times 23!}{23!} = 600$$

Combinations

We observed that the possible permutations of three letters a,b,c were abc, acb, bac, bca, cab and cba. If the order of arrangement is disregarded, all these 6 permutations can be represented by only one combination abc.

The combination of n different objects taken r at time is a selection of r out of the n objects with no attention given to the order of arrangement. The number of combinations of n objects taken r at a time is denoted by

$$^nC_r = \frac{n!}{r!\,(n-r)\,!}.$$

Example 12. From 5 boys and 6 girls find the number of committees of 3 that can be formed with 2 boys and 1 girl.

The number of ways of selecting 2 boys out of 5 is

$$^5C_2 = \frac{5}{2!3!} = 10$$

Similarly one girl out of sic can be selected in

$$^6C_1 = \frac{6!}{1!5!} = 6 \text{ ways.}$$

Using rule 1, the number of committees that can be formed with 2 boys and 1 girl will be

$$10 \times 6 = 60$$

Axiomatic Approach to Probability:

The axiomatic approach to probability closely relates the theory of probability with modern theory of functions and set theory. The auxiomatic approach, proposed by a Russian Mathematician, A.N. Kolmogorov, in the year 1933 includes both the 'Classical' and the 'Statistical' definitions as special cases and also overcomes the limitations in each definition.

For a proper formulation of the approach, we first define some basic terms –

Random Experiment

Any operation that results in two or more outcomes in called an experiment. Here we confine ourselves to such experiments which can be repeated under more or less identical conditions and the results of an individual experiment are unpredictable. For example, an experiment may be conducted for observing the number of accidents in a town, recording the whole sale price of a commodity, observing the daily maximum temperature in a city, and so on.

Sample Space

The set S of all possible outcomes in an experiment is called a sample space. Each element of a sample space is called a sample point. The following examples will clarify.

Example 13. In tossing a fair coin, there are two possible outcomes, namely head *(H)* and tail (1). Thus the sample space is $S = \{H, T\}$

Example 14. When two coins are tossed together, the sample space will be

$$S = \{HH, HT, TH, IT\}$$

Here *HT* represents head on first coin and tail on the second. Similarly, *TH* represents tail on first coin and head on the second. *HH* shows head on each coin and *IT for* tail on each coin. Sometimes it is more convenient to represent the possible outcomes by digits. For example, in the present case, the sample space may be

S= {0, 1, 1,2},

where each digit stands for the number of 'head up' in the experiment.

Example 15. In throwing a die, there are six possible out comes and the sample space is

$$S= \{1,2,3,4,S,6\}$$

Example 16. In a simultaneous throw of two dice, we have 6 x 6 = 36 possible out comes. Thus, the sample IS

S = {(1, I), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, I), (3, 2), (3,3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4,3), (4,4), (4, 5), (4, 6), (5,1), (5, 2), (5, 3), (5, 4), (5,5), (5,6), (6,1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)

In the ordered pair of values, the first number denotes the outcome on the first die and the second on the other die.

Example 17. If two children are selected randomly from group of 3 boys and 2 girls, the sample space in this experiment would be

$S = \{B_1B_2, B_1B_3, B_1G_1, B_1G_2, B_3\, B_2B_3, B_2G_1, B_2G_2\, B_3G_1, B_3G_2, G_1G_2\}$

Here the three boys are designated as B_1, B_2, B_3 and the girls as G_1, G_2.

Example 18. In a simultaneous toss of a die and coin, the sample space will be

S = {(1, *H*), (2, *H*), (3, *H*), (4, *H*), (5, *H*), *(6,H)*, (1,T), (2, T), (3,T), (4, T), (5, T), (6, T)}

Event

Any subset of the sample space is called an event.

Example 9. In a single throw of a die, the event of getting 6 is given by, $E = \{6\}$. Clearly $E \subset \Sigma$.

Example 19. Let S be a sample splice and f be the null set. Since

1. $\phi \subset$ S, so f is an event, called an impossible event.
2. $\Sigma \subset \Sigma$, so S is an event, called a sure event.

Example 20. We observed that S= {1, 2, 3, 4, 5, 6} is the sample space in throwing a die. Let A be the event of getting a number less than 1 and, let B be the event of getting a number less than 7. Obviously, A = ϕ, a no outcome can result in a number less than 1, so A is an impossible event. On the other hand, B = S as level") outcome will be a number less than 7, so B is a sure event.

Simple or Elementary Event

An event containing only a single sample point is called a simple or elementary event.

Example 21. In tossing a coin, the event A *of* getting a tail is a simple event. Here, S = $\{H, T\}$ and A = {T}.

Example 22. Ina simultaneous toss of two coins each one of the events $\{HH\}$, $\{HT\}$, $\{TH\}$, $\{TT\}$ is a simple event.

Compound or Composite or Mixed Events

An event containing more than one sample point is called a compound or composite or mixed event

Example 23. In tossing a fair die, the event A of 'getting an even number', is a compound event.
Here, S= {12,3,4,5,6} and A = {2,4,6}.

Example 24. In a dice-play experiment, the event A of getting a number more than 4" is also a compO1wd event. Here, S= {1,2,3,4,5,6} and A = *{5,6}*

Mutually Exclusive Events

The two events A and B associated with the same experiment are called mutually exclusive if the sunsets of the sample representing the two events are disjoint, i. *e.,* $A \cap B = f$. In other words, no sample point is common to both the events and hence they cannot happen simultaneously. This also agrees with the earlier definition.

Example 25. *Consider the events* .

A = *The number* is *less than* 3

B = *The number* is *more than* 5.

associated with the random experiment of throwing a die. Here the two events A *and* B *are such that both of them cannot happen simultaneously whatever be the outcome of the experiment. Here, the events are* A = {1,2} *and* B= {6}. *Thus,* $A \cap B$= *q" an empty set, i.e., the subsets representing the two events are disjoint, so* A *and* B *are* mutually exclusive.

Complementary Events

In a random experiment, let S be the sample space and let A be an event. Then $\overline{A} \subset S$. Clearly,

$A \subset s$. So $\overline{A}$ is also an event, called the complementary of A. Sometimes, we also denote $\overline{A}$ by A^c or A'. Thus, $\overline{A}$ is the event consisting of all the sample points of the sample space which do not belong *to* A. Obviously A and $\overline{A}$ are mutually exclusive, *i.e.,* $A \cap \overline{A} = f$

Axiomatic Approach or Modern Definition of probability

From the earlier definitions, it is clear that the probability is the limit of the proportion of times that a certain event *A* will occur in repeated trials of an experiment. Obviously, the probability should be a number between zero (0) and unity (I), the probability of the events S (S being the sample space) should be 1, and, for two disjoint events *A* and *B,* the probability of the union of these events should be equal to the sum of their individual probabilities. Thus, these three properties must be satisfied by a probability model. These three properties or requirements are termed as axioms.

We have also seen that event is a set of outcomes, thus the probability of an event is a function defined on set of points. With all these aspects, we define probability as –

Definition

The probability P of an event A with regard to a sample space S of an experiment satisfies the following axioms –

- 0? P(A) ? 1
- P(S) = 1
- For every finite or infinite sequence of disjoint events A_1, A_2, …………

$$P (A_1 \quad A_2 \ \ldots\ldots\ldots) = P (A_1) + P(A_2) + \ldots\ldots\ldots\ldots$$

In a particular, if there are n possible outcomes of a random experiment and all these outcomes are equally likely, then the sample space S will consist of n samples points and probability associated to each sample point would be 1/n [due to axiom (ii)]

Now, if an event A consists of m sample points, then the probability of the event A will be

$$P(A) = m \, X \, \frac{1}{n} = \frac{m}{n}$$

$$= \frac{\text{Number of sample point in A}}{\text{Number of sample point in S}}$$

$$= \frac{\text{Number of cases favorable to event A}}{\text{Number of all possible outcomes in experiment}}$$

= $\underline{n\ (A)}$

Probability of Simple Event

Example 26. Find the probability of getting a tail in a throw of a coin.

Solution: Since there are only two outcomes in tossing a single coin, *i.e.,* Head *(H)* or tail *(T).* Thus, the sample space is $S = \{H,T\}$. The event of going a tail is $E = \{T\}$.

$n(S) = 2$ and $n(E) = 1$

Probability of getting a tail = $P(E) = \frac{n(E)}{n(S)} = \frac{1}{2}$

Example 27. *An unbiased die is thrown. What is the probability that digit 2 appears?*

Solution: In a single throw of a die, there are six possible out comes. Thus $n(S) = 6$. Now let E be the event of getting 2. Then $E = \{2\}$ and $n\{E\} = 1$.

The required probability = $P(E) = \frac{n(E)}{n(S)} = \frac{1}{6}$

Example 28. *In a simultaneous throw of two dice, find the probability of getting a total of 6.*

Solution: In a simultaneous throw of two dice, we have 6x 6 = 36 possible outcomes. Therefore n (S) = 36. Further, let E denotes the event of getting a total of 6.

Then E = {1,5), (2,4), (3,3), (4,1), (5,1) and n (E) = 5

.: The required probability = $P(E) = \frac{n(E)}{n(S)} = \frac{5}{36}$

Example 29. *A coin is successively tossed three times, Find the probability of gelling (i) exactly one head (ii) exactly two heads, (ii) Exactly one head or exactly two heads,*

Solution: In tossing a coin three times, the sample space will be

S = *{TIT, HIT, THT, TTH, THH, HTH, HHT, HHH}* and *n(S)* = 8.

(i) Let *A* denote the event of getting exactly one head. So *A* = *{HIT, THT, TTH}* and *n(A)* = 3,

Thus, the required probability $= P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}$

(ii) Let *B* denote the event of getting exactly two heads, So

B = *{HHT. HTH, THH}* and *n(B)* = 3,

Example 30. *One card is randomly drawn from a pack of 52 cards, Find the probability that*

(i) *the drawn card is red, (ii) the drawn card in an ace, (iii) the drawn card is* red and a king, *the drawn card is* **red or a king.**

Solution: In randomly drawing a card from 52 cards we have 52 possible outcomes. So $n(S) = 52$.

(i) Let *A* denotes the event that the drawn card is red, since the number of red cards is 26,

so $n(A) = 26$.

$$\therefore P(A = \text{a red card}) \quad \frac{n(A)}{n(S)} = \frac{26}{52} = \frac{1}{2}$$

(ii) Let *B* denotes the event that the drawn card in an ace. Since we have four aces, therefore

$n(B) = 4$, .

$$P(C) = \frac{n(C)}{n(S)} = \frac{4}{52} = \frac{1}{13}$$

(iii) Let C denotes the event that the drawn card is red and a king, Since there are only two cards which are red kings. Therefore, $n(C) = 2$.

I

$$P(C) = \frac{n(C)}{n(S)} = \frac{2}{52} = \frac{1}{26}$$

(iv) Let *D* denotes the event that the drawn card in red or a king. Clearly we have 26 red cards, which

include 2 kings and there are two more kings. Therefore $n(D) = 26 + 2 = 28$.

$$\therefore P(D = \text{a red card or a king}) = \frac{28}{52} = \frac{7}{13}$$

Example 31. *A bag contains* 3 *red,* 6 *white and* 7 *blue bal/s. Two bal/s are drawn at random. Find the probability that*

(i) both the bal/s are white,

(ii) both the balls are blue.

(iii) one ball is red and the other is white,

(iv) one is white and the other is blue.

Solution: Total No, of balls in the bag = 3 + 6 + 7 = 16 balls.

Two balls can be drawn out of 16 in $^{16}C_2 = \frac{16 \times 15}{1 \times 2} = 120$ Ways

Therefore, $n(S) = 120$.

(i) Let *A* be the event that both the balls are white. The number of ways of selecting 2 white balls

out of 6 is $^{6}C_2 = \frac{6 \times 5}{1 \times 2}$ Therefore, $n(A) = 15$.

$$p[A = \text{both white balls}] = \frac{n(A)}{n(S)} = \frac{15}{120} = \frac{1}{8}$$

(ii) Let *B* = both the balls are blue. Then the number of ways of selecting 2 blue balls out of 7 is

$^{7}C_2 = \frac{7 \times 6}{1 \times 2} = 21$. Therefore $n(B) = 21$, and

$p[B = \text{both blue balls}] = \frac{21}{40} = \frac{7}{}$

- 40

(iii) Let C = one ball is red and the other is white. *A* red and a white ball can be drawn in $^{3}C_1 \times {}^{6}C_1$

= 18 ways, ∴ n (C) = 18 and

P [C = one red and one white ball] = $\frac{n(C)}{n(S)} = \frac{18}{120} = \frac{3}{20}$

(iv) Let D = one ball is white and the other blue. A white and a blue ball can be drawn in $^{6}C_1 \times {}^{7}C_1$

= 42 ways. Thus $n(D) = 42$ and

$$P[D] = \frac{n(D)}{n(S)} = \frac{42}{120} = \frac{7}{20}$$

Example 32. *A committee of4 people is to be appointed from* 3 *officers of the production department,* 4 *officers of the purchase department, two officers of the sales department, and one chartered accountant. Find the probability of forming the committee* in *the following manner*

(i) *There should be one from each category.*

(ii) *It should have at least one from the purchase department. (iii) The chartered accountant must be* in *the committee.*

Solution: In all the 4 categories, we have 3 + 4 + 2 + 1 = 10 persons and out of them 4 people can be selected in

$$^{10}C_4 = \frac{10 \times 9 \times 8 \times 7}{1 \times 2 \times 3 \times 4} = 210 \text{ ways. Therefore } n(S) = 210.$$

(i) When one from each category is selected: Let *A* denotes the event that a person from each category is selected. *A* person from each category can be selected in $^3C_1 \times {}^4C_1 \times {}^2C_1 \times {}^1C_1 = 24$ ways. Therefore $n(A) = 24$.

(ii) Let *B* denotes the event that the committee consists of at least one from the purchase department.

For getting the needed probability, it is easier to proceed with the complementary event, *i.e., B,* where *B* is the event that no one is selected in the committee from the purchase department. Thus, four persons in the committee are selected from the remaining 3 + 2 + 1 = 6 persons. This, event can occur in $^6C_4 = 15$ ways.

Thus, $n(\overline{B}) = 15$.

$P(\overline{B})$ = no. one is selected from purchase department)

$$= \frac{n(\overline{B})}{n(S)} = \frac{15}{210} = \frac{1}{14}$$

$$P(B) = 1 - P(B) = 1 - \frac{1}{14} = \frac{13}{14}$$

(iii) Let C = one red and one white ball] $= \frac{n(C)}{n(S)} = \frac{18}{120} = \frac{3}{20}$

(iv) Let D = one ball is white and other blue. A white and a blue ball can be drawn in $^6C_1 \times {}^7C_1$ = 18 ways. Thus n(D) = 42 and

$$P[D] = \frac{n(D)}{n(S)} = \frac{42}{120} = \frac{7}{20}$$

Example 33. A committee of 4 people is to be appointed from 3 officer of the production department, 4 officers of the purchase department, two officers of the sales department, and one charted accountant, find the probability of forming the committee in the following manner

- There should be one from each category.
- It should have at least one form the purchase Department.
- The charted accountant must be in the committee.

Solution : In all the 4 categories, we have 3 +4+2+1 = 10 personas and out of them 4

People can be selected in $^{10}C_4 = \frac{10 \times 9 \times 8 \times 7}{1 \times 2 \times 3 \times 4}$ = 210 ways. Therefore n(S) = 210

- When one from each category is selected: Let A denotes the event that a person from each category is selected. A person from each category can be selected in $^3C_1 \times {}^4C_1 \times {}^2C_1 \times {}^1C_1$ = 24 Ways. Therefore n(A) = 24.
- Let B denotes the event that the committee consists of at least one from the purchase department. For getting the needed probability, it is easier to proceed with the complementary event, i.e., B, where B is the event that no one is selected in the committee from the purchase department. This, event can occur in $\overline{^6C_4}$ = 15 ways. Thus, n (B) = 15

$\therefore$ P($\overline{B}$ = no. one is selected from purchase department)

$$\frac{n(\overline{B})}{n(S)} = \frac{15}{210} = \frac{1}{14}$$

- $P(B) = 1 - P(\overline{B}) = 1 - \frac{1}{14} = \frac{13}{14}$

- Let C = one C.A. in the committee of four. One C.A. and three others can be selected $^9C_3 \times {}^1C_1$ = 84 Ways. Therefore, n(C) = 84

$$P(C) = \frac{n(C)}{n(S)} \quad \frac{84}{210} \quad \frac{2}{5}$$

Example 34. *What is the chance that a leap year selected at random contains* 54 *Sundays?*

(52 X 7 + 2 = 366). These extra two days may form the following combinations

(1) Sunday and Monday (2) Monday and Tuesday

(3) Tuesday and Wednesday (4) Wednesday and Thursday

(5) Thursday and Friday (6) Friday and Saturday

(7) Saturday and Sunday. .

Thus, out of the above 7 cases only two, *i.e.,* case (1) and (7) are favorable to the event. Thus, the required probability = 2/7.

Example 35. *A room has* 3 *lamps. From a collection of* 10 *light bulbs a/which* 6 *are no good, a person selects* 3 *at random and puts them in a socket. What* is *the probability that he* will *have light.*

Solution: There are 10 bulbs in all and of these 3 can be selected

in $^{10}C_3 = \frac{10 \times 9 \times 8}{1 \times 2 \times 3} = 20$ ways.

Therefore *n(S)* = 120.

Let *A* be the event of getting light. Then $\overline{A}$ denotes the event of having no light. Now 6 bulbs are not good. Hence the number of ways in which all no good bulbs are chosen is

$^6C_3 = \frac{6X5X4}{1X2X3} = 20.$

Thus $n(\overline{A}) = 20$

$$P(A) = \frac{n(A)}{n(S)} = \frac{20}{120} = \frac{1}{6}$$

$$P(A \text{ having light}) = 1 - P(\overline{A}) = 1 - \frac{1}{6} = \frac{5}{6}$$

Example 36. *A and B throw with* 2 *dice. If A throws 9,find B's chance of throwing a higher number.*

Solution: Here *n(S)* = 6 x 6 = 36. Further, let *A* denotes the event of getting a number higher than 9, *i.e.,* 10, 11 or 12. Then

Some Probability Rules

Often it is easier to compute the probability of an event from known probabilities of other events. This can be well observed if the given event can be represented as the union of two other events or as the complement of an event. Two such rules used for simplifying the computation of probabilities of events are

1. Addition Rule of Probability
2. Multiplication Rule of Probability.

Addition Rule of Probability

For any two events A and B

$$P(A \cup B) = P(A) + P(B) - \mathrm{P}(A \cap B) \qquad \ldots\ldots(8)$$

or $$P(A \text{ or } B) = \mathrm{P}(A) + P(B) - \mathrm{P}(A \text{ and } B) \qquad \ldots\ldots(9)$$

In case A and B arc mutually exclusive events, then $P(A \cap B)$ = a and the addition rule of probability in (8) becomes

$$P(A \cup B) = P(A) + P(B) \qquad \ldots(10)$$

Proof. For any two events A and B, we can write

$$A = (A \cap B) \cup (A \cap \overline{B})$$

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) \qquad \ldots(a)$$

Using axiom (iii) of probability as the events $(A \cap B)$ and $(A \cap B)$ are mutually exclusive.

Similarly –

$$B = (A \cap B) \cup (A \cap \overline{B})$$

$$P(B) = P(A \cap B) + P(\overline{A} \cap B) \qquad \ldots(b)$$

The events $(A \cap B)$ and $(A \cap B)$ being mutually exclusive. Thus, from (a) and (b), one gets

$$P(A) + P(B) == P(A \cap B) + P(A \cap \overline{B}) + P(A \cap B) + P(\overline{A} \cap B) \ldots. \text{(c)}$$

Now the last three terms on R.H.S. of (c), *i.e.,* $P(A \cap B)+ P(A \cap B)+ P(A \cap B)$ represent the probability of occurrence of the events *A* or *B* or both *A* and *B,* i.e., $P(A \cup B)$. Thus, replacing these three terms by $P(A \cup B)$, equation (c) can be written as

$$P(A) + P(B) = P(A \cap B) + P(A \cup B)$$

or $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The rule in *(d)* is called the addition of rule probability. If *A* and *B* are mutually exclusive, $P(A \cap B)$ = a and the addition rule of probability becomes

$$P(A \text{ u } B) = P(A) + P(B) \qquad \text{…. (e)}$$

Example 37. *What* is *the probability of getting an odd number in tossing a die?*

Solution: There are three odd numbers on a die, i. *e.,* 1, 3 and 5. Let *A,* Band C be the respective events of getting 1,3 and 5. *Thus,* $P(A)$ == 1/6, $P(B)$ = 1/6 and $P(C)$ = 1/6. Since *A,* Band C are mutually exclusive, therefore

$$P\,(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}.$$

Actually the Probability Theory provides a means of getting an idea of the likelihood of occurrence of different events resulting from a random experiment in terms of quantitative measures between zero and one. The probability of an impossible event is zero while that of a sure event is unity, Probability can be defined in three ways. According to the Mathematical definition, the probability of an event *A* is

$$P\,(A)\ \frac{m}{n} = \frac{\text{Number of favourable cases to the event } A}{\text{Exhaustive number of cases}}$$

and the probability that the event A does not happen, i.e., $P(\bar{A})$ is

$$P\,(\bar{A}) = 1 - P\,(A) = 1 - \frac{m}{n} = \frac{n-m}{n}$$

However, this definition is not without limitations. In Statistical or Empirical definition, the probability is defined as the limiting value of the number of times the event A happens to the number of trials, as the number of trials becomes infinite. Symbolically

$$P(A) = \lim_{n ? \ ?} \frac{m}{n}$$

However, this definition of probability also has some limitations. According to the axiomatic approach to probability, the probability P of an event A with regard to a sample space S of an experiment satisfies the following axioms -

(a) $0 \ ? \ P(A) \ ? \ 1$

(b) $P(S) = 1$

(c) $P(A) = P(A_1) + P(A_2) + \ldots + P(A_r)$

where A_1, A_2 …... Ar, are the sample points comprising the event A.

There are fundamental rules of probability which are useful for simplifying the calculation of probabilities of mutually exclusive and compound events. These rules are -Additive and Multiplicative rules of probability. According to additive rule, if A and B are two events, then

$$P(A \cup B) = P(A) + P(B) - P(A \ ? \ B)$$

However, if A and B are mutually exclusive, then Additive rule is

$$P(A \cup B) = P(A) + P(B)$$

Further, for two dependent events A and B, the multiplicative rule can be put as

$$\text{P(A r) P } (A \ ? \ B) = P(A).P(B/A)$$

However, if A and B are independent, then the above rule becomes

$$\text{P } (A \ ? \ B) = P(A).P(B)$$

The random variable *is a numerically valued function defined on a sample space of an experiment and assumes different values with a definite probability associated with each value. The distinct values of a random variable X together with their associated probabilities define the probability distribution of the random variable*. Thus, the probability distribution of a random variable can be shown in the form of a table.

x	x_1	x_2	x_3	…	…	…	…	…	x_k
$P(X=x)$	$f(x_1)$	$f(x_2)$	$f(x_3)$	...	...	...	...	...	$f(x_k)$

Now, the Mathematical expectation or simply expectation of the random variable X, denoted as $E(X)$, is

$$E(X) = \overset{k}{\underset{i=1}{?}} x_i f(xi) = x_1 f(x_1) + x_2 f(x_2) + \ldots.. + x_k f(x_k)$$

List of Formulae For Your Information and Understanding

1. **Mathematical Probability**:

If a trial results in n exhaustive mutually exclusive and equally likely cases and m of them and favourable to the happening of an event A, then,

$$P(A) = \frac{m}{n},\ \mathrm{P(A)} = 1 - \frac{m}{n}$$

2. Statistical Probability:

If in n trails, an event A occurs m times, then

$$P(A) = \lim_{n?\ ?} \frac{m}{n}$$

3. **Axiomatic Approach to Probability**

The probability $P(A)$ of an event A with regard to a sample space S of an experiment satisfies the following axioms

(i) $0 ? P(A) ? 1$

(ii) $P(S) = 1$

(iii) $P(A) = P(A_1) + \mathrm{P}(A_2) + \ldots.. + P(A_r)$

4. Important Rules of Probability:

(a) Additive Rule:

$$P (A \cup B) = P (A) + P (B) - P (A \cap B)$$

$$P (A \cup B) = P (A) + P (B) \qquad \text{[if } A \text{ and } B \text{ are mutually exclusive]}$$

(b) Conditional Probability:

$$P (B/A) = \frac{P (A \cap B)}{P (A)}$$

$$P (A/B) = \frac{P (A \cap B)}{P (B)}$$

(c) Multiplicative Rule :

(i) Dependent events.

$$P (A \cap B) = P (A).P (B/A) = P (B).P (A/B)$$

(ii) Independent events.

$$P (A \cap B) = P (A).P (B).$$

5. Mathematical Expectation:

If X is a random variable with probability distribution

x	x_1	x_2	x_3	……	x_k
$f(x)$	$f(x_1)$	$f(x_2)$	$f(x_3)$		$f(x_k)$

Then, the expected value of X is

$$E (X) = \sum_{i=0}^{k} x_i f (x_i) = x_1 f(x_1) + f(x_2) + \quad + x_k f(x_k)$$

PROBABILITY DISTRIBUTION

Objectives

The study of this lesson will help you understand the following objectives:

- Understand the probability distribution
- Understand the properties of Probability distribution
- How it is used in business activities.

Hello, in the previous classes we discussed "**observed frequency distributions**" which are results or outcomes of actual observations and experimentation. For example, data related to marks, height, weight and age of students may be classified in the form of an observed frequency distribution. However, instead of observed frequencies in theoretical probability distributions there are all possible values of a random variable and the frequencies are replaced by actual probabilities, which depend on the nature of a random variable. For example, if X is a random variable showing the numerical value of the outcome in an experiment of rolling a six faced die. Then X has six possible outcomes 1, 2, 3, 4, 5 or 6 and to each outcome there is an associated probability 1/6, which we calculate on the basis of theoretical considerations that all the outcomes are equally likely. Thus, the theoretical probability distribution of the numbers when a die is tossed becomes

Theoretical Probability Distribution of the numbers when a die is tossed.

Outcome (x)	1	2	3	4	5	6	Total
Probability of (x) = $P(x)$	1/6	1/6	1/6	1/6	1/6	1/6	1.0

A few other examples of such theoretical probability distributions have already been discussed in the previous chapter. It is observed that many random variables associated with statistical experiments have similar properties or have the same general type of behavior and therefore can be described by the same probability distribution. For example, in the case of die tossing experiment, the theoretical probability distribution can be described as –

$P(x) = 1/6$; x = 1, 2, 3, 4, 5 and 6.

Now suppose the experiment of tossing a die is repeated 60 times, then the expected or theoretical frequency distribution of x, denoted by $f_e(x)$ will be:

Theoretical Frequency Distribution of the numbers when a dice is tossed

x	1	2	3	4	5	6	Total
$P(x)$	1/6	1/6	1/6	1/6	1/6	1/6	
$f_e(x) = N.P(x)$	10	10	10	10	10	10	$N = 60$

Thus, a theoretical probability distribution of x can be used to define an expected or theoretical frequency distribution of x as

$F_e(x) = N.P(x)$

Next, if we consider that the same experiment of die tossing experiment is empirically repeated 60 times, then the observed frequency distribution of x may be of the following form

Observed Frequency Distribution of the numbers when a dice is empirically tossed 60 times

x	1	2	3	4	5	6	Total
Observed frequency $f_o(x)$	12	8	11	9	8	12	60

Comparing the above theoretical and observed frequency distributions we conclude that theoretical distributions may not fully agree with the observed or empirical distributions yet it is likely that the observed distributions would become closer and closer to the theoretical distributions when the number of repeating an experiment is large. In view of this, theoretical probability or frequency distributions can be used for representing an observed probability or frequency distribution.

Again since we know that a random variable may either be discrete or continuous. Thus, the distributions of discrete and continuous random variables are accordingly called **discrete or continuous probability distributions.**

There are a large number of theoretical or expected probability distributions in statistics, however, in the present chapter, only three such distributions have been discussed. These distributions are

1. Binomial distribution
2. Poisson distribution
3. Normal distribution.

Binomial Probability Distribution

Suppose there is an experiment consisting of repeated independent trials, each with two possible outcomes, 'Success or Failure'. For example, in a coin tossing experiment we have only two outcomes, head or tail, which may also be called as 'success' or failure'. This is also true if 7 cards are drawn in succession from a pack of cards and each trial labeled as 'success' or failure', depending on whether the card is red or black. In this card drawing experiment, if each drawn card is replaced before the next drawing, then the repeated trials are independent and probability of success in each trial remains constant as 1/2 from trial to trial. Experiment of this nature is called Bernoulli trial. *Thus*, the Bernoulli trials *have the following properties*

1. Each trial has only two possible outcomes, ‘Success’ or ‘Failure’.
2. The repeated trials are independent.
3. The probability of success in each trial remains constant.

Binomial probability distribution has been developed to find the probability of x successes in n Bernoulli’s trials. In this regard, let us suppose that

1. The experiment consists of *n* repeated trials.
2. In a trial, the occurrence of an event be considered as ‘success’ and nonoccurrence as ‘failure’. Let *p* be the probability of ‘success’ and $q = (1-p)$ be the probability of failure in single trial.
3. Since *x* denotes the number of successes in n independent repeated trials, therefore *x* is a random variable which can take any of the values 0, 1, 2, ... ,*n*.

With these notations, let us first consider the probability of *x* successes and (*n x*) failures in a specific order. Here each success occurs with probability *p* and each failure with probability *q*. Thus, the probability of *x* success and (*n x*) failures in the specified order can be obtained by using multiplicative rule of probability as under –

$$\underbrace{p.p.p \quad p}_{x \text{ times}} \, . \underbrace{q.q.q \quad q}_{(n\ x) \text{ times.}}$$

$$= p^x q^{(n-x)}$$

But we are interested in any *x* trials resulting in success and these *x* trials out of *n* can be chosen in ^nC_x mutually exclusive ways of ordering them. Thus, on adding the probabilities of all ^nC_x cases or simply multiplying $p^x q^{(n-x)}$ by ^nC_x we get the general formula of computing the probability of *x* successes in *n* Bernoulli trials as –

$$P(x) = p^x q^{(n-x)} + p^x q^{(n-x)} + \dots + p^x q^{(n-x)} \ (^nC_x \text{ times})$$

$$= {}^nC_x \, p^x q^{(n-x)}$$

where $x = 0, 1, 2, ..., n$.

Now, by putting each value of the random variable *x* in (2), you get the binomial probability distribution as shown in the following table

Binomial Probability Distribution

No. of Success (x)	Probability Function $P(x) = {}^nC_x p^x q^{(n-x)}$
0	$^nC_0 p^0 q^n = q^n$
1	$^nC_1 p \ q^{n-1}$
2	$^nC_2 p^2 \ q^{n-2}$
3	$^nC_3 p^3 \ q^{n-3}$
:	:
x	$^nC_x p^x q^{(n-x)}$
:	:
n - 1	$^nC_{(n-1)} p^{n-1} q$
n	$^nC_n p^n q^0 = p^n$
Total	$(q + p)^n = 1$

The probability distribution in the above table is named as binomial probability distribution. It is named from the fact that the $(n + 1)$ terms in the binomial expansion of the term $(q + p)^n$ respectively correspond to the values of $P(x)$ in (2) for $x = 0, 1, 2, ..., n$. The binomial expansion of the term $(q + p)^n$ is

$$(q + p)^n = {}^nC_0 p^0 q^n + {}^nC_1 p^1 q^{(n-1)} + {}^nC_n p^n q^0$$

$$= P(0) + P(1) + P(2) + \dots + P(n)$$

In the above discussion, we observed that the formula in (2) represents all the probabilities of the random variable x and therefore is called the probability function of the binomial probability distribution. From (3), it is clear that sum of all the probabilities is unity, i.e.

$$\sum_{x=0}^{n} P(x) = P(0) + P(1) + P(2) + \dots + P(n)$$

$$= {}^nC_0 p^0 q^n + {}^nC_1 p^1 q^{(n-1)} + {}^nC_2 p^2 q^{n-2} + \dots + {}^nC_n p^n q^0$$

Binomial Frequency Distributions

In case of binomial probability distribution, let us suppose that n trials constitute an experiment and if this experiment is repeated *N* times, then on using formula in (1), the theoretical or expected binomial distribution is given by

$$f(x) = N\, P(x) = N\,{}^nC_x p^x q^{(n-x)};\ x = 0, 1, 2, \dots n$$

On varying *x* from 0 to *n*, the expected binomial frequency distribution can be obtained as shown in the following table.

Binomial Frequency Distribution

x	$f(x) = N\,P(x) = N\,^nC_x\,p^x q^{(n-x)}$
0	$f(0) = Nq^n$
1	$f(1) = N\,^nc_1\,pq^{n-1}$
2 :	$f(2) = N\,^nc_2\,pq^{n-2}$ ::
x :	$f(x) = N\,^nC_x\,p^x q^{(n-x)}$:
n	f (n) Np^n
Total	$N(q+p)^n = N$

Properties of the Binomial Probability Distribution

1. *n* and *p* are the two parameters of the binomial distribution. As soon as the values of *n* and *p* are known, the binomial distribution is completely determined.
2. The mean of the binomial distribution is *np*.
3. The standard deviation and the variance of the binomial distribution are Ö(*npq*) and (*npq*) respectively. Since $q < 1$, therefore $np > npq$. Thus, the mean of the binomial distribution is always greater than its variance.
4. lf $p = q = 1/2$, then the binomial distribution is a symmetrical distribution.
5. For *p* ¹ *q*, the binomial distribution is a skewed distribution.
6. Binomial distribution is a discrete probability distribution.

Important Remark

The probability function of the binomial distribution [equation (2)] is used in computing the probability of x successes in n bernoulli trials. For writing the probability function, one should first ascertain the value of p, the probability of success in a trial. The following examples will clarify the concept.

Example 1. The mean and variance of a binomial distribution are 2.5 and 1.875 respectively. Obtain the binomial probability distribution.

Solution: Since the mean and variance of the binomial distribution are NP and npq respectively. Thus np = 2.5 and npq = 1.875.

$$\frac{npq}{np} = \frac{1.875}{2.5} = q = 0.75$$

p=1- q =1- 0.75 = 0.25

Again since, np = 2.5 or n. (0.25) = 2.5

$$n = \frac{2.5}{0.25} = 10$$

Now, with n = 10, p = 0.25 and q = 0.75, the binomial probability distribution is

$$p(x) = {}^{10}C_x\,(0.25)^x\,(0.75)^{10-x},\ x = 0,1,2,\ldots 10$$

Where x is the number of successes.

Example2. Comment on the following statement-

For a binomial distribution, mean = 8 and variance =10.

Solution: We know that the variance of the binomial distribution is always less than its mean. But in the given statement variance = 10 > mean = 8.Thus, the statement is incorrect because, for a binomial distribution if we consider

Mean = np = 8 and variance = npq = 10

Then $\frac{npq}{np} = \frac{10}{8}$ gives q = 1.25 > 1.

Since q is the probability of failure which cannot exceed 1, the given statement is not correct.

Example 3. If mean and variance of a binomial distribution are 4 and 2 respectively. Find the probability of (i) exactly two successes (ii) less than two successes (iv) At least two successes.

Solution: given that np= 4 and npq=2

q=½, p= ½ and n = 8

Thus, the binomial probability distribution giving the probability of x successes in 8 trial is $p(x) = {}^8C_x(½)^x(½)^{8-x}$

$= {}^8C_x\,(½)^8$, x = 0, 1, 2…8 ……(a)

1. For obtaining the probability of exactly 2 successes, we put x=2 in (a) and get

   $$P(2) = {}^8C2(½)^8, =28.(½)^8 = \frac{28}{256} = \frac{7}{64}$$

2. Here we need the probability of less than two successes.i.e.,

   p(x < 2) = p(x = 0) + p(x = 1) = p(0)+p(1)

   $$= {}^8C_0(½)^8 + {}^8C_1(½)^8 = \frac{1}{256} + \frac{28}{256} = \frac{29}{256}$$

3. The required probability is

   p(x > 6) = p(7) + p(8)

   $$= {}^8C_7(½)^8 + {}^8C_8(½)^8 = \frac{8}{256} + \frac{1}{256} = \frac{1}{256}$$

4. p (At least two sucesses) = P(x ³ 2) = p(2)+ p(3)+….+ p(8)

$$= \sum_{x=2} p(x) = \sum_{x=0} p(x) - \sum_{x=0} p(x)$$

$$= 1 - [\,p(0) + p(1)\,]\ \left[\text{As} \sum_{x=0}^{8} p(x) = 1\right]$$

$$= 1 - \left[\frac{29}{256}\right] = \frac{227.}{256}$$

Example 4. A binomial variable x satisfies the relation 9 p(x=4) = p(x=2) when n=6. Find the value of the parameter p.

[C.A. (Inter) May 1992]

Solution: Since, the binomial probability distribution is

$p(x) = {}^nC_x\,(p)^x(q)^{n-x}$, x = 0,1,2,…n …….(a)

for n = 6, equation (a) becomes

$p(x) = {}^6C_x p^x q^{6-x}$, x=0,1,2,…6. …….(b)

Considering the given relation 9p(x=2), we have

$$\frac{P(x=4)}{P(x=2)} = \frac{1}{9} \quad \text{or} \quad \frac{{}^6C_4 p^4 q^{6-4}}{{}^6C_2 p^2 q^{6-2}} = \frac{1}{9}$$

$$\text{or} \quad \frac{15.p^4.q^2}{15.p^2.q^4} = \frac{1}{9} \quad \text{or} \quad \frac{p^2}{q^2} = \frac{1}{9} \quad \text{or} \quad \frac{p}{q} = \frac{1}{3}$$

$$\text{or} \quad 3p = (1 - p) \quad \text{or} \quad 4p=1$$

Thus, $p = ¼$.

Example 5. In tossing 10 coins. What is the probability of having exactly 3 heads?

Solution: In this example the number of trials made is n=10. Further, the occurrence of head on one coin be taken as success. Then the probability of getting a head in one trial is p=½ and so q = 1 – p = 1 - ½ = ½. Thus, the probability of x successes in n=10 trials is given by

$P(x) = {}^{10}C_x (½)^x (½)^{10-x}$, x=0,1,2,…10.

Here, we need the probability of exactly 3 heads (successes). So putting x = 3 in (i), we get

$$P(3)={}^{10}C_3(½)^3 (½)^7 = {}^{10}C_3(½)^{10} = \frac{10.9.8}{1.2.3} (½)^{10}$$

$$= \frac{120}{1024} = \frac{15}{128}$$

Example6. An ordinary six-faced die is thrown 4 times. What are the probabilities of getting 4,3,2,1,0 aces? Ace means getting number 1 on a die.

Solution: Let getting an ace in a single throw be considered as success. Then p=1/6 so q=1 – p = 5/6. Also given n = 4. Thus, probability of x successes in 4 trials can be obtained by using the binomial probability distribution, i.e.,

$P(x) = {}^4C_x(1/6) (5/6)^{4-x}$, x=0,1,2,3,4.

$P(x=0) = {}^4C_0(1/6)^x(5/6)^4 = (5/6)4$

$P(x=1) = {}^4C_1(1/6)^1(5/6)^3 = 4.1/6(5/6)^3 =125/324$

$P(x=2) = {}^4C_2(1/6)^2(5/6)^2 = 6.?(1/6)^2(5/6)^2 =25/216$

$P(x=3) = {}^4C_3(1/6)^3(5/6) = 4.(1/6)^3(5/6) =5/324$

$P(x=4) = {}^4C_4(1/6)^4 = (1/6)^4$.

Example 7. A coin is tossed six times. What is the probability of getting four or more heads.

Solution: Here n=6,p=possibility of success.(head on one coin)=½

Thus the probability of x successes in 6 trials can be obtained by using the binomial probability distribution

$P(x) = {}^6C_x(1/2)^x(1/2)^{6-x}$

$= {}^6C_x(1/2)^6$, x=0,1,2,….6.

Thus the probability of getting four or more heads. i.e,

$P(x ³ 4) = p(x = 4) + p(x = 5)+p(x = 6)$

$= {}^6C_4 (½)^6 + {}^6C_5 (½)^6 + {}^6C_6 (½)^6 = ({}^6C_4 + {}^6C_5 + {}^6C_6) (½)^6$

$$= (15+6+1)(½)^6 = \frac{22}{64} = \frac{11}{32}$$

Example 8. Five coins are tossed 3.200 times: find the frequencies of the distribution of heads and tails and tabulate the results.

Solution: Here we are given that-

N=3200,n=5,p=½ and q=½

Thus, the binomial probability distribution is

$p(x) = {}^5C_x (½)^x (½)^{5-x}$,

$= {}^5C_x (½)^5$; x=0,1,2,3,4,5 …….(i)

Also, the expected frequency distribution is

$F(x) =Np(x)= N^5.C_x (1/2)5$, x=0,1,2…5. ……(ii)

Putting x=0,1,2,3,4 and 5 in (ii), the binomial frequency distribution is shown in the table.

Binomial frequency distribution

No. of heads x	$p(x) = {}^5Cx (1/2)^5$	F(x) = Np(x) = 3200.p(x)
0	${}^5C_0.1/32 = 1/32$	100
1	${}^5C_1.1/32 = 5/32$	500
2	${}^5C_2.1/32 = 10/32$	1000
3	${}^5C_3.1/32 = 10/32$	1000
4	${}^5C_4.1/32 = 5/32$	500
5	${}^5C_5.1/32 = 1/32$	100
Total	$\sum_{x=0}^{5} p(x) = 1$	N = 3200

Poisson Probability Distribution

Poisson distribution is another discrete probability distribution having specific uses. The distribution was developed by a famous French mathematician Simeon D. Poisson in 1837. For understanding the utility of this distribution let us consider the probability function of the binomial distribution as

$P(x) = {}^nC_x p^x q^{n-x}$; x=0,1,2,…n.

If n = 4 and p = ½, then we can easily calculate probability so 3 successes as

$P(3) = {}^4C_3(½)_3(1/2) = 4. (½)^4 =1/4$

However, if n = 1000 and p = 1/100, then

$P(3) = {}^{1000}C_3(1/100)^3(99/100)^{1000-3}$

But it will be very difficult to calculate the probability of 3 successes in this case. In this regard we come across an interesting property of the distribution. It can be shown that if n is large, p is small and np = m (a constant), then the binomial distribution

$P(x) = {}^nC_x p^x q^{(n-x)}$; x = 0,1,2,…n

Tends to Poisson distribution with probability function

$$P(x) = \frac{e^{-m}m^x}{x!} ; x = 0,1,2,…$$

Thus, when n is large, p is small and np = m (a constant), the limiting form of the binomial distribution is known as Poisson distribution.

Definition of the Poisson distribution

A random variable X taking values 0,1,2,3… is said to have Poisson distribution with parameter m if for (m > 0), the probability of x successes is given by-

$P(x) = \frac{e^{-m}m^x}{x!}$ x = 0,1,2,……

Also S p(x) = $\frac{e^{-m}m^x}{x!}$

$= e^{-m} \left(I = \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!}\right)$

$= e^{-m}.e^{m} = e^0 = 1.$

As assumed, the probability of one success or event is small, therefore Poisson distribution in (5) is also known as the distribution of rare events. In view of this, there is a very wide range of phenomena in various fields where Poisson distribution has specific applications. For example, the behavior of the number of defectives in painted surface of unit area, the number of telephone calls received at particular switch board, the number of deaths due to specific disease etc. can be analyzed by using the Poisson probability distribution. In all such cases probability of occurrence of an event (p) is very small while the number of repeated trials (n) is very large. Putting various values of x=0,1,2,….in(5), the Poisson probability distribution can also be shown in a tabular form as below.

Poisson Probability Distribution

x	0	1	2	3	….
$P(x) = \frac{e^{-m}m^x}{x!}$	e^{-m}	$e^{-m}.m$	$\frac{e^{-m}m^2}{2!}$	$\frac{e^{-m}m^3}{3!}$	….

Poisson Frequency Distribution

As discussed earlier suppose in a Poisson distribution, n trials constitute an experiment and if the experiment is repeated n times, then the theoretical or expected Poisson distribution is given by

$F(x) = N.P.(x) = \frac{N.e^{-m} m^x}{x!}$; x=0,1,2……

On varying x from 0 to 8, the expected Poisson frequency distribution can be obtained in the following tabular form-

Poisson Frequency Distribution

1. There is only one parameter of the Poisson distribution, which is m. Thus, if m is known, the Poisson probability distribution is completely known. If n and p are known, then m=np.
2. The mean of the Poisson distribution is m.
3. The variance of the Poisson distribution is also m. Thus, the mean and the variance of the Poisson distribution are the same.
4. Poisson distribution is also a discrete probability distribution,
5. It is a limiting case of the binomial probability distribution.
6. With the following three conditions, the binomial distribution tends to the Poisson distribution-

(i) p is small (ii)n is large (iii)np=m (a finite constant)

Example 9. If a random variable x follows Poisson distribution much that p(X=1)=p(X=2), then find (I) then mean of distribution (ii)p(X=0).

Solution: the probability function of the Poisson distribution is

$P(x) = \frac{e^{-m}m^x}{x!}$; x=0,1,2,…..

Since p(x = 1) = p(x = 2)

Or $\frac{e^{-m}m}{1!} = \frac{e^{-m}m^2}{2!}$ Or m=2.

Thus, (i) the mean of the distribution is m = 2 in (a), the Poisson probability distribution is

$P(x) = \frac{e^{-m}2}{x!}$; x=0,1,2,…..

$P(x = 0) = e^{-2} = 0.1353$

Example 10. Find the probability that at most 5 defective fuses will be found in a box of 200 fuses if experience show that 2 percent of such fuses are defective.

Solution: Here the probability of a fuse being defective (success)is

P=2/100 = 0.02 and n=200. Therefore, m = n = np = 200x 2/ 100 = 4. Since p is small so the probability of x defective can be obtained by using the Poisson probability distribution, i.e,

$P(x) = \frac{e^{-m}m^x}{x!} = \frac{e^{-4}4^x}{x!}$; x=0,1,2,….

The needed probability is-

P (at the most five defective fuses) = p(x ? 5)

$= P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$

$= e^{-4} + e^{-4}\,4 + \frac{e^{-4}\,4^2}{2!} + \frac{e^{-4}\,4^2}{3!} + \frac{e^{-4}\,4^4}{4!} + \frac{e^{-4}4^5}{5!}$

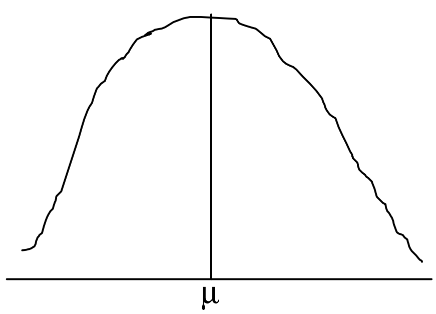
$= e\text{-}4\left(1 + 4 + 8 + \frac{4^3}{6} + \frac{4^4}{24} + \frac{4^5}{120}\right)$

$= 0.0183\left(1 + 4 + 8 + \frac{4^3}{6} + \frac{4^4}{25} + \frac{4^5}{120}\right) = 0.785$

The Normal Probability Distribution

So far we have discussed tow discrete distributions, namely, the binomial and the Poisson, these distributions enable us to find the probabilities of distinct events, like the probability of defective items in a sample of given size the probability of accidents in a factory, In general, with these distributions, we are able to enumerate the probability of successes for failures occurring in a fixed number of independent trails. However in practice, we come across a number of biological, social, economic, industrial and psychological measurements where the variables are continuous in nature, and as such can be adequately described only by a continuous probability distribution, One of the most important continuous probability distributions in the entire field of statistics is the normal probability distribution. It has been observed that a vast number of variables arising in studies of natural social psychological and economic phenomena confirm t o normal distribution, the graphical shape of the normal distribution, called the normal curve, is the bell shaped smooth symmetrical curve as shown in the figure below. You just see to it.

Generally, the distributions of quantitative data show concentration of frequencies near the central value of the distributing and then the frequencies gradually taper off symmetrically on both sides of central value. This general tendency of data, for a very large number of observations, give rise to the symmetrical bell shaped form of normal curve, Thus the normal curve or distribution is a theoretical model which may be used to describe the frequency distribution of a vast variety of continuous variables,



The Normal Curve

Definition

A continuous random variable x it said to be normally distributed if it has the probability density function represented by the equation-

$$P(x) = \frac{1}{\sigma\sqrt{(2\pi)}^2} - \left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]; -\infty < x < \infty \quad \ldots\ldots.(7)$$

Where, μ = mean of the normal distribution

σ = Standard devoting of the normal distribution

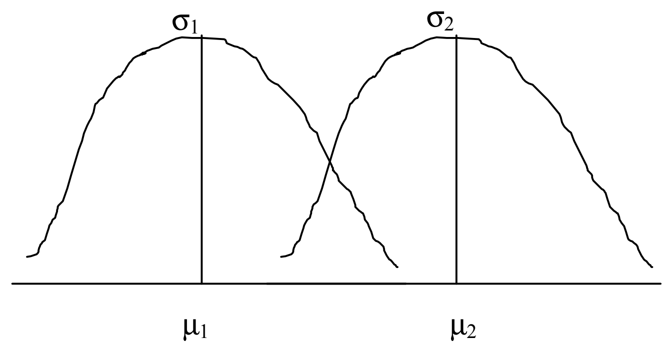
π = 3.14159 approximately.

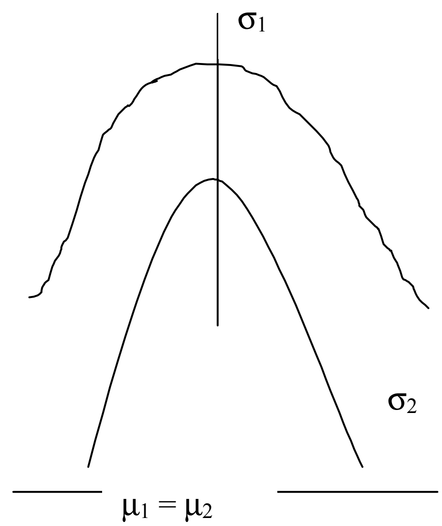
ε = 2.70828 approximately.

m and σ are also known as the two parameters of the normal distribution, once the values of m and σ are known the shape for the equation of the normal distribution is completely determined. The idea will be more clear from the shapes of some normal curves for different values of m and σ.

This figure shows the graphs of two normal curves with unequal means (μ_1 ¹ μ_2)

But equal standard,

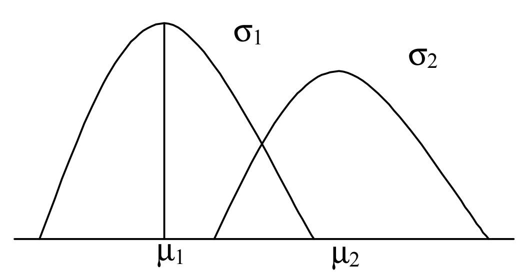




Two nominal Curves with μ_1 '" μ_2 and $\sigma_1 = \sigma_2$ Two normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$ are shown above.

We have two normal curves with equal means ($\mu_1 = \mu_2$) but different standard deviations (σ_1 ¹ σ_2). Here the two curves are centered over the same point, but the curve with smaller standard deviation is higher and narrower in range.

Finally fig.4 shows the sketches of two normal curves with different means and different standard deviations.



Properties of the Normal Distribution or Curve

Some of the important properties of normal distribution or curve may be listed as under-

1. The curve is symmetrical about the vertical axis the two halves of the curve would coincide.
2. As a result of symmetry, the mean median and mode of the distribution are identical .i.e.,

Mean = Median= Mode

3. Since there is only one maximum point in the curve the normal curve is unimodel, i.e it has only one mode.
4. μ and σ respectively denote the mean and the S.D. of the distribution . These (m and s) are also known as two parameters of the distribution.
5. The curve is asymptotic to x-axis i.e it becomes closer and closer to x-axis but never actually touches it
6. The mean deviation= 4/5 standard deviation.
7. The height of the curve declines symmetrically in either direction from the maximum point. Therefore, the ordinates for values of $x = \mu \pm k$, where k is a real number, are equal. For example,

The heights of the curve or ordinates at $x = \mu + \sigma$ and $x = \mu - \sigma$ are exactly the same. This is also clear from fig.5.

8. The total area under the normal curve and above the horizontal axis is 1.0, which is essential for a probability distribution or a curve.

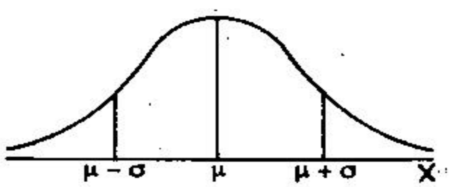


Fig. 5. Ordinates of a normal curve at two equidistant values from mean.

9. One of the most important properties of the normal curve is the area property. Since the shape of the normal curve is complete determined by its parameters m and e, the area under the curve bonded by the two ordinates also depends on these parameters. Some important areas under the curve bounded by the ordinates at e, 2σ, and 3σ distances away from mean in either direction are shown in fig. 6, 7 and 8 respectively. The Fig. 6 shows that the area between ordinates at $x = \mu - \sigma$ and $x = \mu + \sigma$ is 0.6827 or 68.27%.

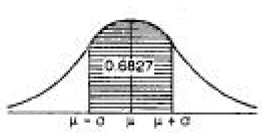


Fig. 6. Area Between 1σ limits (68.27%)

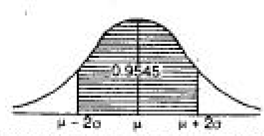


Fig. 7. Area between 2σ limits (95.45%)

In Fig. 7, the shaded portion is the area between ordinates at $x = \mu - 2s$ and $x = \mu + 2\sigma$. The area is 0.9545 or 95.45%.

Also the area between ordinates at $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$ is 0.9973, i.e., the area under the normal curve beyond these ordinates is only 1-0.9973 = 0.0027 which is very small. Thus, practically, we can say that the whole area under the normal curve lies with in $\mu \pm 3\sigma$ limits, which are also called 3-sigma limits. In probability terms, these areas can be summarized as –

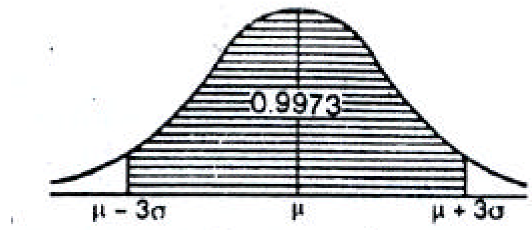


Fig. 8. Area between 3σ limits (99.73%)

1. Area within 1s limits = $P[\mu - \sigma < x < \mu + \sigma] = 0.6827$
2. Area within 2s limits = $P[\mu - 2\sigma < x < \mu + 2\sigma] = 0.9545$
3. Area within 3s limits = $P[\mu - 3\sigma < x < \mu + 3\sigma] = 0.9973$

10. It is possible to transform any normal random variable X with mean m and variance s^2 to a new normal random variable Z with mean 0 and variance 1. This normal random variable Z with mean 0 and variance 1 is called standard normal variable (S.N.V.). The transformation of X to Z is

$$Z = \underline{X - \mu} \qquad \ldots\ldots(8)$$

Area Under the Normal Curve

The area or probability under a normal probability distribution or curve bounded by two ordinates at x=a and x=b is written as – $P[a \le x \le b]$

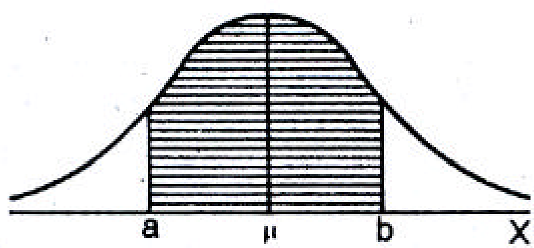


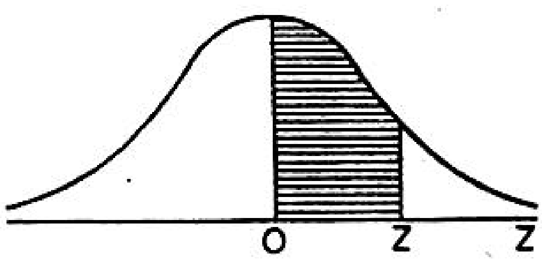
Fig. 9. Shaded area is = $P[a \leqslant x \leqslant b]$

This probability in (9) is the probability that a normally distributed variable x lies between two specified values a and b and can be represented by the shaded area in Fig. 9. Given below. Further it is notable that the area or probability under a normal curve depends on its parameters m and s. Thus, the area under the normal curve will change with the values of m and s. Fortunately, it is possible to transform any normal random variable X to standard normal variable Z by using the transformation –

$$Z = \frac{X - \mu}{\sigma}$$

The area, proportion or probability under the standard normal distribution or curve between the ordinates at mean 0 and Z has been given in Table A.(Given

In the end of the book). The values of Z range from 0 to 3.99. Thus, Table A can be used to measure area under standard normal curve between ordinates at mean 0 and at some positive value of Z as shown by the shaded area in Fig.10 Thus, for studying the area under a normal curve, we have to follow the following steps-



1. Use the transformation in (8), u,w, Z= X- μ/σ for converting the given normal random variable X to a standard normal variableZ.
2. Thus the probability under the normal curve between the ordinates at x=a and x=b, i.e $p(a \le x \le b)$will correspond to the area under standard normal curve between the ordinates

Z1= a - μ/σ and Z2=μ - u/σ . Symbolically, we can write-

$P(a \le x \le b) = p(a - \mu/\sigma \le x-\mu/\sigma \le b-\mu/\sigma)$

$= p(Z1 \le Z \le Z2)$ (Using the transformation $z = \frac{x-\mu}{\sigma}$)

Hence, $Z1 = \frac{a-u}{\sigma}$ and $Z1 = \frac{b-u}{\sigma}$

This correspondence in probability is also shown in Fig.11

3. As a third step we use table A to compute the needed probability or area, i.e

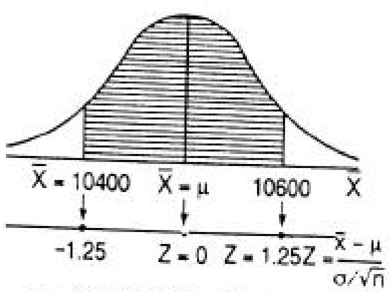


Fig. 31. $P(10,400 < \bar{x} < 10,600) = P(-1.25 < Z < 1.15)$

$$p\left(\frac{a-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right)$$
$$= p(Z1 \le Z \le Z2)$$

Note that value of μ, σ, a and b are known

How to Use Table A

The following important points should be kept in mind while computing area or probability under a standard normal curve-

1. The total area under the standard normal curve is 1.
2. The mean of the distribution is 0(zero). Thus, the negative and positive values of Z will lie on the left and right of mean respectively.
3. The ordinate at mean, i.e at Z=0 divides the area under the standard normal curve into two equal parts. Thus the area on the right and left of the ordinate at Z=0 is 0.5. Symbolically,

$$P(-\infty \le Z \le 0) = P(-a \le Z \le \infty) = 0.5$$

4. Since the curve is symmetrical, thus-
   $P(0 \le Z \le a) = P(-a \le Z \le 0)$.
   This will be more clear from Fig.12
   The following examples will clarify the procedure discussed.

$$P[-a < Z < 0] = P[0 < Z < a]$$

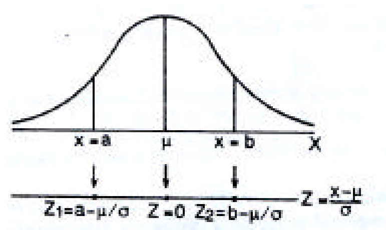


Fig. 12 In view of the symmetry of the curve

Example 14. X is a normal variable with the mean μ=25 and standard deviation = 5,

Find the values of Z_1 and Z_2 such that

$P(20 \le X \le 30) = p(Z1 < Z < Z2)$

Here, Z is a standard normal variable.

Solution: For trans forming a normal variable X to a standard normal variable Z, WE use the transformation in equation (8), i.e'

$$Z = \frac{X-\mu}{\sigma} = \frac{X-25}{5}$$

Thus, for values of X = 20 and X = 30, the corresponding values of Z variable will be

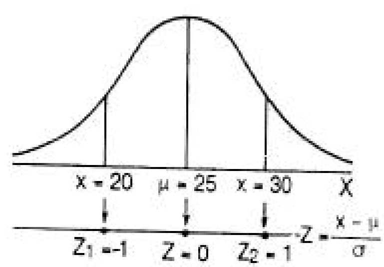
$Z_1 = \frac{X-\mu}{\sigma} = \frac{20-25}{5} = -1$ (when X = 20)

And $Z_2 = \frac{X-\mu}{\sigma} = \frac{30-25}{5} = +1$ (when X = 30)

There fore,

$P(20 < X < 30) = p(-1 < Z < 1)$

The transformation from X to Z values is also shown graphically in Fig.13.



Transformation from X to Z

Example 15 Z is a standard normal variable. Use table A to determine the following probabilities-

(i) $p(0 < z < 1.2)$ (ii) $p(-1.2 < z < 0)$ (iii) $P(-2 < Z < 2)$.

Solution : For determining the probability or area under a standard normal variable, we use table A given in the appendix.

1. The probability expression P(0<Z<1.2) is the area between the ordinates at Z=0 and Z=12,thus for Z=1.2 the required probability is directly determined from table A , as

$P(0 < Z < 1.2) = 0.3849$.

2. For determining p(-1.2 < Z < 0) , we should recall that in table A we are given area of probability between ordinates at mean Z=0 and Z, where Z is some positive value, For determining probabilities other than directly given in table A, We as a rule should always make use of the symmetrical properties of a normal curve to convert the required area in the form given in table A .Table A is then used for determining the required area or probability . Thus

   p(-1.2 < Z < 0) = p (Z < 0 < 1.2)= 0.3849.

3. For computing p(-2 < Z < 2) , let us sketch the area as shown in fig.14 . The shaded area is p(-2 < Z < 2).

Thus mathematically we can determine the area as under .

P(-2 < Z < 2) = P(-2 < Z < 0) + P(0 < Z < 2)

= P(0 Z 2) + P(0 Z 2)

[As P(-2 < Z < 0) = P(0 < Z < 2) by symmetry]

= 2P(0 < Z < 2) = 2(0.2772)

= 0.9544.

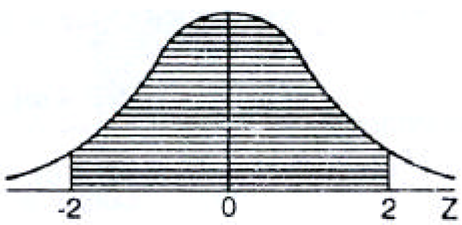


Fig. 14. Shaded area is P(– 2 < Z < 2).

Example 16. If Z is a standard normal variable. Find the following probabilities or areas

(i) P (-2 < Z < 1)   (ii) P(-∞ < Z > -1)

(iii) P (2 < Z < ∞)   (iv) P(-∞ < Z < 1)

Solution: (i) A rough sketch of the needed area, as shown by shaded area in Fig. 15, is helpful in its determination.

From this figure, we observe that

P(-2 < Z < 1) = P(-2 < Z < 0) + P(0 < Z < 1)

= P(0 < Z < 2) + P(0 < Z < 1)

= 0.4772 + 0.3413

[For Z = 2.0 and 1.0 from table A]

= 0.8185.

1. P(-∞ < Z < -1) is represented by the shaded area in
2. Fig. 16

   P[-∞ < Z < -1]

   = P[-∞ < Z < 0] – P[-1 < Z < 0]

   = P[0 < Z < 8] – P[0 < Z < 1]

   = 0.5000 – 0.3413

   = 0.1587.

3. P(2 < Z < ¥) is the shaded area in Fig. 17.

   Thus, P(2 < Z < ∞)

   = P(0 < Z < ∞) – P(0 < Z < 2)

   = 0.5000 – 0.4772 (From table A)

   = 0.0228.

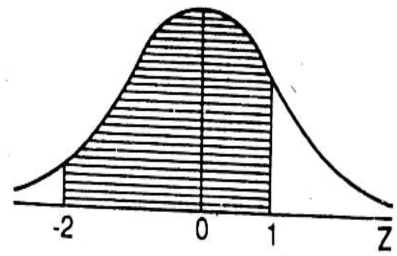


Fig. 15. Shaded area P (– 2 < Z < 1)

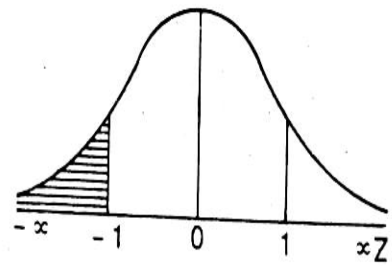


Fig. 16. P (–∞ < Z < – 1) = Shaded area

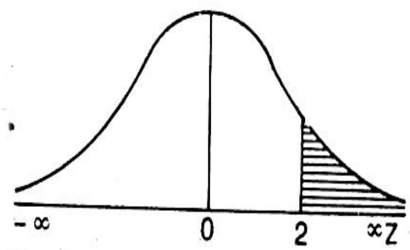


Fig. 17.. Shaded area = P (2 < Z < ∞)

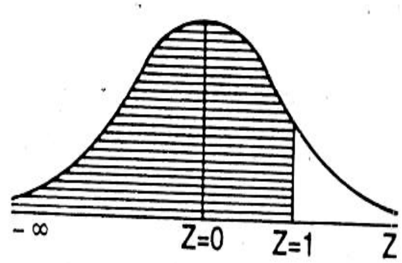


Fig. 18.. Shaded area = P (–∞ < Z < 1)

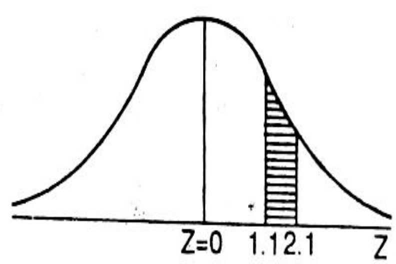


Fig. 19.. Shaded area = P (1.1 < Z < 2.1)

4. The shaded area in Fig. 18 represents $P(-\infty < Z < 1)$. Thus,

$P(-\infty < Z < 1)$

$= P(-\infty < Z < 0) + P(0 < Z < 1)$

$= P(0 < Z < \infty) + P(0\ \ Z\ \ 1)$

$= 0.5000 + 0.3413$ (From Table A)

$= 0.8413.$

Example 17. For a standard variable Z, find

i. $P(1.1 < Z < 2.1)$

ii. $P(-2.0 < Z < 1.2)$

Solution: (i) Shaded area in Fig. 19 represent $P(1.1 < Z < 2.1)$

Thus,

$P(1.1. < Z < 2.1)$

$= P(0 < Z < 2.1) - P(0 < Z < 1.1)$

$= 0.4821 – 0.3643$

$= 0.1178$

i. Shaded area in fig. 20 is $(-2.0 < Z < -1.2)$. Therefore,

$P(-2.0 < Z < -1.2)$

$= P(-2.0 < Z < 0) - P(-1.2 < Z < 0)$

$= P(0 < Z < 2) - P(0 < Z < -1.2)$

$= 0.4772 – 0.3849 = 0.0923.$

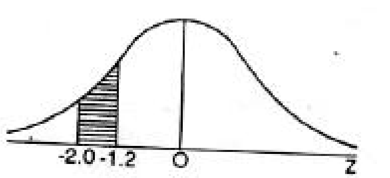


Fig. 20. Shaded area $= P(-2.0 < Z < -1.2)$

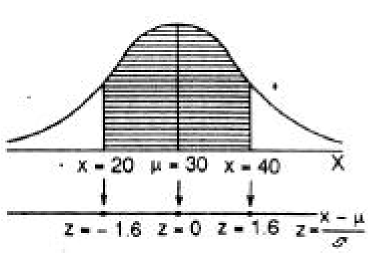
[By symmetry]

Example 18. 2000 students appeared in an examination. Distribution of marks is assumed to be normal with mean m = 30 and e 6.25. How many students are expected to get marks.

1. Between 20 and 40 2. Less than 35.

Solution: Let X stands for the marks of the students. Thus, X is a normal variable with mean m = 30 and e = 6.25. Thus,

$$Z = \frac{X - m}{\sigma} = \frac{X - 30}{6.25}$$



1. In the first case, we need to find the area or probability of X lying between ordinates at X = 20 and X = 40, i.e.,

$$P(20\ \ X\ \ 40) = P\left(\frac{20-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{40-\mu}{\sigma}\right)$$

(Using $Z = \frac{X-\mu}{\sigma}$)

$= P\left(\frac{20-30}{6.25} < Z < \frac{40-30}{6.25}\right)$

$= P(-1.60 < Z < 1.60)$

(This probability is the shaded area in Fig. 21)

$= P(-1.60 < Z < 0) + (0 < Z < 1.60)$

$= P(0 < Z < 1.60) + P(0 < Z < 1.60)$

(By symmetry of the normal curve)

$= 2.P(0 < Z < 1.60)$

(Using table A for Z = 1.60)

$= 2\ (0.4452)$

$= 0.8904$ or 89.04%

Therefore, out of 2000 students, the expected number of students scoring between 20 and 40 us 2000 x $P(20 < X < 40)$

= 2000 x 0.8904

=1780.8 = 1781.

2. In the second case we need to calculate the $P(X < 35)$. For its computation we first transform the variable X to Z by using the transformation $Z = \frac{X-\mu}{\sigma}$ as

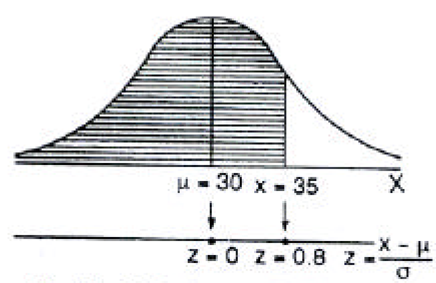


Fig. 22. Shaded area $= P(X < 35)$
$= P(Z < 0.80)$

[Using the area sketch in Fig. 22]

$$P(X < 35) = P\left(\frac{X-\mu}{\sigma} < \frac{35-\mu}{\sigma}\right)$$

$= P\left(\frac{X-30}{6.25} < \frac{35-30}{6.25}\right) = P(Z < 0.80)$

$= P(-\infty < Z < 0) + P(0 < Z < 0.80)$

$= P(0 < Z < \infty) + P(0 < Z < 0.80)$

$= 0.5000 + 0.2881$ (Using table A)

$= 0.7881$

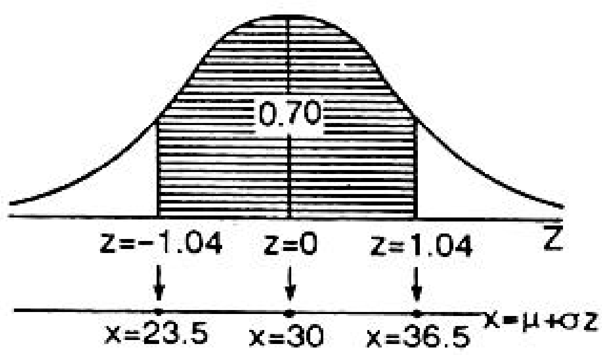
Thus, the expected number of students scoring less than 35 marks will be

2000 x $P(X < 35)$ = 2000 x 0.7881

= 1576.2 @ 1576

Example 19. In the above example 18, determine the central limits of scores within which 70 percent of the students score.

Solution: In the previous example, we were required to compute the area of probability under the normal curve bounded by ordinates at different values of Z, but the present problem is just the reverse. In this case, we have to find the centrally located limits, s –a to a, within which 70 percent or0.70 area (0.035 on each side of mean) is covered as shown in fig.23.



As we know, the transforming the variable from X to Z is

$\frac{X-\mu}{\sigma}$ Here we first . The same

transformation or relationship can also be used for transforming the variable Z to X in the following manner-

$X = \mu + \sigma Z$

Now proceeding as in the example, we are given that

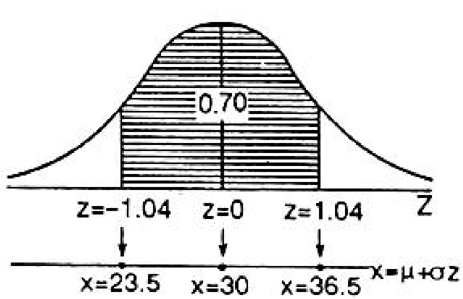
$P(-a < Z < a)=0.70$

or $P(-a < Z < 0) + P(0 < Z < a) = 0.70$

or $P(0 < Z < a) + P(0 < Z < a)=0.70$

or $2P(0 < Z < a) = 0.70$

or $P(0< Z < a) = 0.35$ ——(ii)



Now we go through the area in Table A and locate a value of Z=a corresponding to the area 0.35 as in equation (11) above. Such a value is Z=a=1.04.Therefore,the centrally located limits on Z scale covering 0.70 area are -.04 to 1.04.finally,using the relationship in (1),the two limits on Z scale can be transformed on to X scale as under-

Putting Z=-1.04 in (i) , one gets the lower limits on X scale as-

$X = \mu + \sigma Z = 30 + 6.25 (-1.04) = 30 = 6.50$

$=23.5 = 24$ in whole number

Similarly For Z = 1.04 one gets the upper limits on X scale as-

$X = \mu + \sigma Z = 30 + 6.25(-1.04) 30=6.50$

$= 36.5 = 37$ in whole numbers.

Therefore, the central limits of scores with in which 70 percent of students score are 24 to 37.

Example 20. Assuming the mean height of students is an exactly normal distribution be 68.22 inches and variance of 10.8 inches squares. How many students in a college of 1000 students would you expect to be over six feet tally?

Solution: let X be the variable representing height of students, Then X is a normal variable with mean m = 68.22 inches and variance = 10.8 inches squares. The transformation from X to Z is Z = X-μ /s. Here we first compute the probability p(x 72) as over 6 feet means X > 72 inches.

Now considering

$$P(X > 72) = p\left(\frac{X-\mu}{\sigma} > \frac{72 - 68.22}{\sqrt{10.8}}\right)$$

$$= P(Z > 1.15)$$

$$= P(0 < Z < \infty) - P(0 < Z < 1.15)$$

$$=0.5000 - 0.3749$$

$$= 0.1251$$

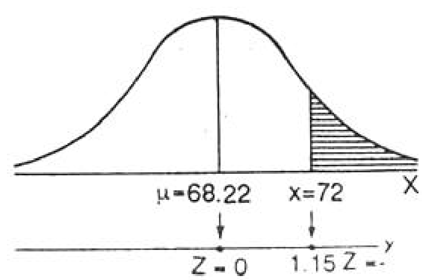


Fig. 24. Shaded area = $P(X > 72)$
= $P(Z > 1.15)$

Thus. The expected number of students having height more than 6 feet or 72 inches will be

1000 x P(X > 72) = 1000 x 0.1251 = 125.1 @ 125

Example 21. In a distribution exactly normal 7% of the items are under 35 and 89% are under 63. What are the mean and standard deviation of the distribution?

Solution: Let X be the normal variable with mean u and standard derivation. Then we are

given that

$P(x < 35)=0.07$and

$P(X < 63) = .89$

The locations of the points X = 35

$Z = \frac{35-\mu}{\sigma} = -Z_1$(say)

And for X = 63

$Z = \frac{63 - \mu}{\sigma} = Z_2$(say)

(Z2 will be positive as it lies to the right of mean)

Using the given probabilities in (i) and (ii) and their sketches in Fig.25 we can easily see that

$P(0 < Z < Z_1) = .43$ and $p(0 < Z > Z_2)=.39$

From table A in the appendix we get

$Z_1= 1.48$ and $Z_2 =1.23$

Thus from (iii) and (iv) one gets

$$\frac{35 - \mu}{\sigma} = - 1.48 \quad \text{and} \quad \frac{63 - \mu}{\sigma} = 1.23$$

Subtracting one gets, $\frac{28}{\sigma} = 2.71$ or $s = 10.33$

And m = 35+1.48x10.33=50.3

Thus the mean and standard deviation of the normal distribution are 50.3 and 10.33 respectively.

Example 22. Of a large group of men 5% are under 60 inches in height and 40% are between 60 and 65 inches. Assuming a normal distribution, find the mean height and standard deviation.

Solution: Let x be the normal variable with mean u and standard deviations. Then we are given that

P(X<60) = 0.05and p(60<X65) = .40. From the sketch of the given probabilities as shown in fig.26. It is clear that both the points X=60 and X=65 are located to the left of then mean and so the corresponding values of Z will be negative.

When X= 60 , $Z = \frac{X- \mu}{\sigma} = \frac{60 - \mu}{\sigma} = -Z_1$(say)

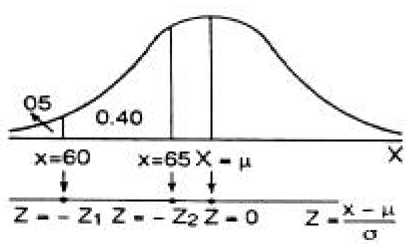


Fig. 26. *P* (*X* < 60) = 0.05 and *P* (60 < *X* < 65) = 0.40.

and $X = 65$, $Z = \frac{X - m}{\sigma} = \frac{65 - m}{\sigma} = - Z_2$ (say) ……(ii)

From Fig. 26, it is clear that

$P(-Z_1 < Z < 0) = P(0 < Z < Z_1) = 0.45$

And $P(-Z_2 < Z < 0) = P(0 < Z < Z_2) = 0.05$

Therefore, from Table A,

$Z_1 = 1.645$ and $Z_2 = 0.13$

Thus, from (i) and (ii), one gets

$$\frac{60 - \mu}{\sigma} = -1.645$$

and

$$\frac{65 - \mu}{\sigma} = -0.13$$

solving (v) and (vi), we get

$\mu = 65.42$ and $\sigma = 3.29$

Example 23. The local authorities in a certain city install 10,000 electric lamps in the streets of the city. If these lamps have an average life of 1000 burning hours with a standard deviation of 200 hours, assuming normality, what number of lamps might be expected to fail (i) in the first 800 burning hours? (ii) Between 800 and 1200 burning hours? After what period of burning hours would you expect what (a) 10% of the lamps would fail? (b) 10% of the bulbs would be still burning?

Solution: Let X denote the life of the bulbs in burning hours. Then X is normally distributed with mean m = 1000 and S.D. = e = 200

i. Here, we first calculate –

$P(X < 800) = P \left(\frac{X - \mu}{\sigma} < \frac{800 - \mu}{\sigma}\right)$ [using Z = X – m /s]

$= P\left(\frac{X - 1000}{200} < \frac{800 - 1000}{200}\right)$

$= P(Z < -1)$

$= P(-\infty < Z < 0) - P(-1 < Z < 0)$ [see the standard area (lined) in Fig. 27]

$= P(0 < Z < \infty) - P(0 < Z < 1)$ [using Table A]

$= 0.5000 - 0.3413 = 0.1587$

Thus, out of 10,000 lumps, the number of lumps failing before 800 hours is

10,000 x P(X < 800) = 10,000 x 0.1587 = 1587

ii. In this case we consider –

$P(800 < X < 1200)$

$= P\left(\frac{800 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{1200 - \mu}{\sigma}\right)$

$= P\left(\frac{800 - 1000}{200} < Z < \frac{1200 - 1000}{200}\right)$

$= P[-1 < Z < 1]$

$= P[-1 < Z < 0] + P[0 < Z < 1]$

$= P[0 < Z < 1] + P[0 < Z < 1]$

$= 2P[0 < Z < 1] = 2.0 \times 0.3413$

$=0.6826$

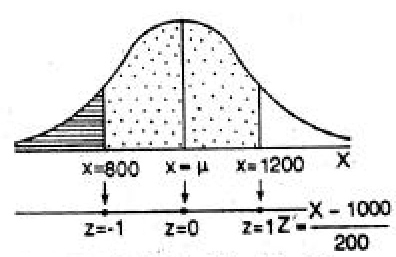


Fig. 27. *P* (*X* < 800) = Lined shaded area. *P* (800 < *X* < 1200) = Dotted area

[Using Table A]

Thus, the expected number of lamps with life between 800 and 1200 hours is –

10,000 x P(800 < x < 1200) = 10,000 x 0.6826 = 6825

a. Let 10% of the lamp fail after x1 hours of burning period. Then we have to find the value of x1 such that

$P(X < x_1) = 0.10.$

Or on transforming on Z scales, we have for X = x1,

$Z = \frac{x_1 - \mu}{\sigma} = \frac{x_1 - 1000}{200} = -Z_1$ (say)

Z_1 will be negative as it lies to the left of mean (see sketch in Fig. 28).

Now $P(Z < -Z_1) = 0.10$ or $P(Z < Z_1) = 0.10$

Or $P(0 < Z < \infty) - P(0 < Z < Z_1) = 0.10$

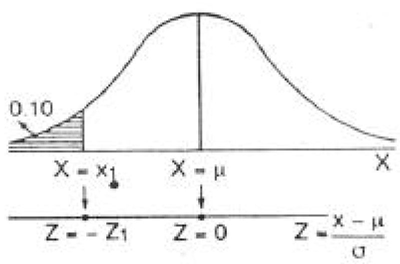


Fig. 28. $P(X < x_1) = 0.10.$

Or $0.5000 - P(0 < Z < Z_1) = 0.10$

Or $P(0 < Z < Z_1) = 0.5000 - 0.10$

$= 0.40$

From (i) $-Z_1 = 1.28 = \frac{x_1 - 1000}{200}$

And $x_1 = 1000 - 1.28(200) = 744$ hours.

Thus, after 744 hours we expect 10% of the lumps to fail

$P(X > x_1) = 1.10$

Let 10% of the lumps are still burning after x1 hours of burning period. Then we are given that

$P(X > x_1) = 0.10$

Since x_1 lies on the right of mean so the corresponding Z value will be positive, i.e.,

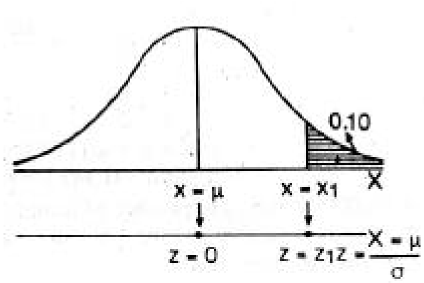
When $X = x_1$, $Z = \frac{x_1 - 1000}{200} = Z_1$ (say)

Considering $P(X > x_1) = P\left(\frac{x - \mu}{\sigma} > \frac{x_1 - \mu}{\sigma}\right) = 0.10$

Or $P(Z > Z_1) = 0.10$

Or $P(0 < Z < \infty) - P(0 < Z < Z_1) = 0.10$

Or $0.5000 - P(0 < Z < Z_1) = 0.10$



Or $P(0 < Z < Z_1) = 0.40$

$Z_1 = 1.28$ [From Table A in the appendix]

$1.28 = Z_1 = \frac{x_1 - 1000}{200}$

or $x_1 = 100 + 200 \times 1.28 = 1256$ yours.

Thus we expect that 10% of the bulbs will be still burning after 1256 hours of burning life.

Example 24. The weekly wages of 2000 workers in a factory are normally distributed with a mean of Rs. 200 and a variance of Rs. 400. Estimate the lowest weekly wages of the 197 highest paid workers and the highest weekly wages of the 197 lowest paid workers

Solution : Lit X be the variable denoting weekly wages. Thus X is normally distributed with mean m = 200 Rs. And standard deviation s = Ö(400) = 20 Rs. Now, let x_2 denotes the lowest weakly wages of the 197 highest paid workers, i.e 197 workers get weakly wages more than Rs.x_2 or symbolically

$P(X > x_2) = 197.2000=0.0985$

Or $P\left(\frac{X - \mu}{\sigma} > \frac{x_2 - \mu}{\sigma}\right)=0.0985$

Or $P(Z > \frac{x_2 - 200}{200}) = 0.0985$

Or $P(Z > Z_2) = 0.0985$ where $Z_2 = \frac{x_2 - 200}{200}$ ……(i)

Here the value of Z_2 will be positive as the corresponding area is on the right tail of the distribution as shown in fig.30

Now since

$P(Z>Z_2) = 0.0985$

Or $P(0 < Z < \infty) - P(0 < Z < Z_2)=0.0985$

Or $.5000 - P(0 < Z < Z_2) = 0.0985$

Or $P(0 < Z < Z2) = 0.5000-0.0985 = .4015$

Z2 =1.29 (From Table A)

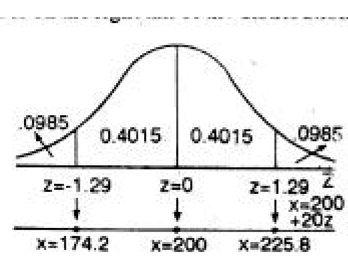


Fig. 30. $P(z > 1.29) = 0.0985$
$P(z < -1.29)$

Thus on using(i)

$Z_2 = \frac{x_2 - 200}{20} = 1.29$

Or $x_2 = 200 + 20 \times 1.29 = 225.80$ Rs.

Thus Rs.225.80 is the lowest wage of the 197 highest paid workers.

Similarly. For computing the highest weekly wages say x 1of 197 lowest paid workers we can write

$$P(X < x_1) = \frac{197}{2000} = 0.0985$$

$$\text{Or } P\left(\frac{X-\mu}{\sigma} < \frac{x_1-\mu}{\sigma}\right) = 0.0985 \text{ or } P\left(Z < \frac{x_1-200}{20}\right)=0.0985$$

Or $p(Z < Z_1) = 0.0985$ where $Z_1=(x_1\text{-}200/20)$

Clearly in this case Z_1 will be negative as the area lies on the left tail of the distribution (seeFig.30.) Considering again the probability

$P(Z < Z_1)=0.0985$

Or $p(-\infty < Z < 0)$- $P(-Z_1 < Z < 0)=0.0985$

Or $P(0 < Z < \infty) - P(0 < Z < Z_1)=0.0985$

Or $.5000\text{-}P(0 < Z < Z1) = 0.4015$

$Z_1=1.29$

from table A.

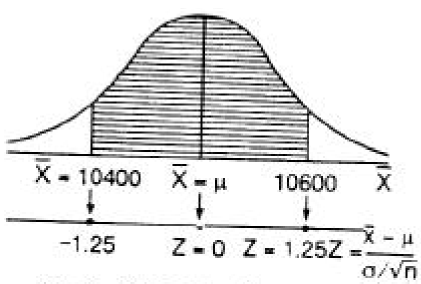


Fig. 31. $P(10,400 < \bar{x} < 10,600)$
$= P(-1.25 < Z < 1.15)$

But since the area is less than 0.5 on he left tail of the distribution, the value of Z1 will be taken as negative.

So, from (ii)

$$Z_1=-1.29 = \frac{x_1-200}{20} \text{ or } x_1= 200\text{-}20 \text{ x}1.29=174.2\text{Rs.}$$

Thus , Rs. 174.2 is the weekly wages of the 197 lowest paid workers.

Important Result

If x is the mean of a random space of size n from a normal population with mean m and variance s^2, then the distribution of the sample mean is a normal distribution with mean m and variance s^2/n.

Example25. A random sample size 100 is selected from a large group of wage corners, The wages are normally distributed with mean annual income or Rs 10,500 with a standard deviation of Rs.800. Find the probability that the sample mean income falls between Rs.10400and Rs.10,600.

Solution: Let x denote the annual wages. Then X is a normally distributed random variable with mean m=10,500 and standard deviation σ = 800 . thus, the sample mean x will be normally distributed with mean μ =10,500 and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{800}}{\sqrt{100}} = 80$

$$\text{Thus,} \quad Z = \frac{x-\mu}{\sigma\sqrt{n}} = \frac{x-10,500}{80}$$

is standard normal variable. Now let us consider the required probability, i.e.,

$P(1,400 < x < 10,600)$

$= P\left(\frac{10400-10500}{80} < \frac{x-10500}{80} < \frac{10600-10500}{80}\right)$

$= P(-1.25 < Z < 1.25)$ [using (i)]

$= P(-1.25 < Z < 0) + P(0 < Z < 1.25)$

$= P(0<Z<1.25) + P(0<Z<1.25)$

(See sketch of Fig. 31)

$= 2P(0<Z<1.25) = 2 \times 0.3944 = 0.7888.$

Thus, the probability that the sample mean x lies between Rs. 10,400 and Rs. 10,600 is 0.7888.

Importance of the Normal distribution

The normal distribution was obtained by a French mathematician De-Moivre in 1733. De Moivre showed mathematically that as n ®¥ and neither p nor q is small, the binomial distribution will tend towards the normal. Thus, like Poisson distribution, normal distribution is also a limiting form of the binomial distribution. More so, Poisson distribution also tends to normal distribution when m, the parameter of the Poisson distribution, increases indefinitely. Further, we observe that not only binomial and Poisson, but many other probability distributions have normal distributions as their limiting forms. Also a large number of sampling distributions tend to normal as the sample size n increases. In view of this, normal distribution plays a central role in statistical analysis. Starting from descriptive statistics to statistical inference, this distribution has a wide range of applications in all areas of statistics.

Relation between Distributions

Binomial and Poisson

If n is large, p is small and np = m (a positive finite constant), then binomial distribution tends to Poisson distribution.

Binomial and Normal

If n is large and neither p nor q is small, then the discrete binomial distribution tends to the continuous normal distribution with mean np and standard deviation Ö(npq).

Poisson and Normal

The Poisson distribution also tends to normal distribution when its parameter m increases indefinitely.

Numerical Problems

Binomial Distribution

1. Comment on the following:

   For a binomial distribution, mean = 8 and variance = 12.

2. For a binomial distribution, the mean is 6 and the standard deviation is /2. Write the binomial distribution.

3. In a binomial distribution with 6 independent trials the probabilities of 3 and 4 distribution.

4. Determine the binomial distribution for which the mean is 4 and standard deviation is

3. A coin is tossed six times. What is the probability of obtaining four or more heads?

5. The incidence of occupational disease in an industry is such that the workers have a 20% chance of surrendering from it. What is the probability that out of six workmen 4 more will contract disease?

6. Find the probability of obtaining exactly three doubles in four throws of a pair of dice.

7. Eight coins are tossed simultaneously. Find the probability of getting at least 6 heads.

8. In a box containing 100 bulbs, 10 are defective. What is the probability that out of a sample of 5 bulbs.

(i) None is defective (ii) Exactly 2 are defective?

Poisson Distribution

9. If the proportion of defectives in a bulk is 4%, find the probability of not more than 2 defectives in a sample of 10. Given that e-0.4=0.6703

10. Find the probability that exactly 2 defectives will be found in a packet of 100 blades if experience shows that 3% of such blade are defective. Given that e-3=0.04978.

11. A car hire from has two cars which it hires out day by day. The number of demands for a car on each day is distributed as Poisson distribution with mean 1.5. Calculate the proportion of days on which (i) Neither car is used (ii) some demand is refused. Given that e-15=0.2231. [Hint.(i) Neither car is used means , no demand , i.e. ,x=0 (ii) some demand is refused . means z>2)=p{x>2} =1-[p (0) +p (1) +p (2)].

12. Criticize the following statement: "The mean of the Poisson distribution is 5 while its standard deviations are 4."

13. (a) If x follows poison law such that p(x=1)=p(x=2), find the mean and variance, Also find p(x=4)

Normal Distribution

14. Z is a standard normal variable. Find the following areas or probabilities: (I)P(0<Z<2) (ii) P(1<Z<2) (iii) P(Z>2) (iv)P(Z>0) (v)P(Z<0) (vi)P(Z<-1.5) (vii) P(-1.5<Z<-1)

15. If Z is a standard normal variable. Find the following probabilities:

(i) P(Z<1.2) (ii) P(Z<-1.2) (iii) P(-1.2<Z<-1.3)

24. If X is normal variable with mean u=30 and variance = 16. Find Z values corresponding to X= 40, 30, and 20.

25. If X is a normal variable with mean u=20 and =10.find P (15<X$0).

26. The mean life of bulbs is normally distributed with mean 120 days and standard deviation 20 days. If 1000 such bulbs are installed then find how many bulbs will fail in less than 90 days.

27. In an intelligence test administered to 1000 students the average score was 42 and standard deviation 24.Find (I) the expected number of students scoring more than 50

(iii) The number of students scoring between 30 and 54. (Then value of score exceeded by top 100 students.

28. Given a normal distribution with mean u=50 and =10. Find then probability that X assumes a value between 45and 62.

29. A certain type of storage battery lasts on the average 30days with a standard deviation of 5 days. Assuming that the battery lives are normally distributed find that the given battery will last less than 23 days.

30. An electrical firm manufactures light bulbs which have a life length that is normally distributed with mean 800 hours and standard deviation of 40 hours. Find the probability that the bulb burns between 778 and 834 hours.

31. Given a normal distribution with u=200 and = 100.Find (i) the area below 214 (ii) the two points containing the middle 75% of the area.

32. Given the normally distributed variable X with mean 18 and standard deviation 2.5 find : (a) The value of c such that p(X<c)=0.2578 (b)The value of k such that p(X<K)=0.1539.

True or False

1. The probability distribution is the outcome of different probabilities taken by the random variable.
2. The mean of the binomial distribution is greater than its variance.
3. The mean and variance of a Poisson distribution are equal.
4. M is the only parameter in a Poisson distribution.
5. Poisson and normal distribution are limiting form of the binomial distribution.
6. The binomial distribution is completely determined if n and p are known.
7. The binomial distribution is completely determined if n and p are known.
8. If n is large m p is small and np=m (a positive number), the binomial distribution tends to normal distribution.
9. If n is large and neither p nor q is small, the discrete binomial distribution tends to the normal distribution.
10. Poisson distribution is used when n is large and p is very small.
11. If p= q, the binomial distribution is symmetrical distribution.
12. The mean of the binomial distribution is 10 and its variance is 12.
13. The mean and variance of the Poisson distribution are 7and 9 respectively.
14. The normal curve is the graphical shape of the normal distribution.

15. The normal curve may be asymmetrical.
16. The normal distribution is completely determined if the value of u and o2 are known.
17. P (-3<Z<3) = 0.9973, if Z is standard normal variable.
18. P (u-3 <X<u+3) = 0.9973, if X is normal variable with mean u and variance .
19. The mean, median and mode coincide for the normal distribution.
20. The shape of the normal curve is determined by its parameters u and .
21. The total area under the normal curve is unity.
22. Binomial distribution is continuous distribution.

STOCHASTIC MODEL

Objectives

On completion of this lesson you would be able to fulfill the following objectives:

- Understanding of the probabilities of moving from one state to another and how it is analyzed
- Identify of how the Markov process can be used in the area of finance.

Dear friends, today we will see to the problems of how stochastic model is important in the area of financial management. To be very realistic, we are very much interested to know as to how a random variable changes over time. And here as a student of finance you may be more interested in the behavior of the price of shares during a certain period of time. In such kind of situations we can use the stochastic model/ process also called Markov chain/process.

Few assumptions underling Markov Analysis

The Markov chain analysis to follow is based on the following assumptions:

1. **Finite states:** The given system has a finite number of states, none of which is "absorbing" in nature. In our detergent example, there are three states- three brands of detergent which the customers can switch between. They are all non-absorbing. To understand this, let us consider what an absorbing state is. In our example, if there was one brand of detergent from which a customer would never switch, then that brand would have represented an absorbing state. Similarly, in the gambler's ruin example discussed earlier, once the person involved reaches the state where he has Rs 40 (having doubled his capital) or Re 0 (having vanished all his capital), he would not play more, and quit, Each of these two, therefore, represents and absorbing state. The analysis to follow assumes that all the states are non-absorbing. We shall later relax this assumption.

In regard to the classification of states, it may further be noted that:

i. For two states i and j, a sequence of transitions that begins in i and ends in j is called a path from i to j.

ii. A state of j is known as reachable from state i if there is a path leading from i to j.

iii. Two states i an j are said to communicate if each is reachable from the other: j is reachable from i, and i from j.

iv. A state i is known as a transient state if there exists a state j which is reachable from i, but i is not reachable from state j. Thus, a state i is transient if there is a way to leave state i that never returns to state i.

v. If a state is not transient, it is known as a recurrent state.

Figure 1 gives the transition from one state to another for the detergent problem in the graphical form. It is evident that each of the three states is "reachable" from all others, and hence communicative. None of these is absorbing and each one is recurrent.

Similarly, Figure 2 depicts the inter-state transition for the gambler's ruin problem. Here, states 0 and 4 are absorbing states, as mentioned earlier, and state 1, 2 and 3 are transient states. For instance, it is possible to go from state 2 to state 4 along the path 2-3-4, there is no way to return to state 2 from state 4.

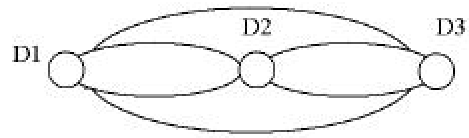


Fig. 1


Fig. 2

2. **First-order process:** The condition (or state) of the system in any given period is dependent only on its condition prevailing in the previous period and the transition probabilities. In our brand switching case, it is assumed that the choice of a particular brand of detergent is dependent upon and influenced only by the choice in the previous month. Similarly, in the gambler's ruin problem, the amount of money that the person in person in question has after t+1 plays of the game depends on the past history of the game only though the amount of money he has after t plays.

   In this context, it may be mentioned that where the probability of the next event depends upon the outcome of the last event, like a customer's choice of detergent brand in a given month depends on the choice in the last month, the Markov process is termed as a first-order Markov process assumes order Markov process assumes that the customers choice next month may depend upon their choices during the immediately preceding two months. Similarly, a third order process is based on the assumption that the customers' choices of a discussion of first order Markov processes only.

3. **Stationarity:** The transition probabilities are constant over time. As already indicated, it is assumed that the system has settled down so that the switching among different brands takes place at the given rates in each time period. Needless to

add, this makes the Markov chain here stationary. With the rules of the game unchanged over time for the gambler's ruin problem, the problem also is an example of stationary Markov chains.

4. **Uniform time periods:** The changes from one state to another take place only once during each time period, and the time periods are equal in duration. Thus, in our example, it is assumed that the customers change their brands of detergent on a monthly basis and accordingly, the monitoring is also done is also done on a month-to-month basis. If it is believed that the customers change the brands within a shorter period, then we may reduce the time periods to, say, weeks.

Analysis: Input and Output

In the Markovian analysis, the analysis of a given system is based on the following two sets of input data – the transition matrix (containing the transition probabilities) and the initial condition in which the system is. Based on these inputs, the model provides for the following predications:

a. The probability of the system being in a given state at a given state at a given future time.

b. The steady state (that is, long run or equilibrium) probabilities.

We shall consider first the inputs and then the analysis and the output.

Inputs

The two inputs, viz. transition probabilities and the initial conditions can be discussed as:

1. **Transition probabilities :** The transition probabilities are required for obtaining both the types of predictions mentioned above. It may be recalled that the Markov process describes movement of the system from certain state in the current stage (may be current time period) to one of the n possible states in the next stage. This movement is in an uncertain environment but we are given the probability associated with any move. This probability is system known as the transition probability and expressed as pij, being the probability that he moves from current state i to another state j in the next time period.

2. **The initial conditions:** The initial conditions describe the situation the system presently is in. For instance, as indicated earlier, as the market is divided 30% , 45% and 25% between the brands D1, D2 and D3 respectively on March 1, the current date, it describes the initial conditions. It may be expressed in terms of a row vector [0.30 0.45 0.25]. In case the initial condition is described as [0 0 1] for the market, it implies that the brand D3. Further, for the gambler's ruin problem, the initial condition is givenby [0 0 1 0 0], which implies that he currently is in the state where his capital is Rs 20.

Output

As stated earlier, there are two predictions, which a Markov analysis provides. The first of these is the probability of the system being in a particular state at a suture time., while the other is the steady state probabilities.

1. **Specific-state probabilities:** For calculating the probabilities for the system in specific states, we let q1(k) to represent the probability (q) of the system being in a certain state (i) in a certain period (k), called the state probability. Since the system would accupy one and only one state at a given point in time, it is obvious that the sum of all q1 values would be equal to 1. In general terms, with a total of n states,

   q1(k) + q2(k) + q3(k) +..........qn(k) =1, for every k

   in which k is the number of transitions (0, 1, 2,..........)

   Let us consider the calculation of the q1(k) probabilities for the detergent example. With states of the system designated as D1, D2 and D3, qD1 (0) represents the probability of the customer choosing brand D1 this month (at t =0) and qD1 (1) represents the probability of choosing this brand after one tr4ansition, that is, the next month. Similarly, qD1 (2) is the probability of choosing this brand after two transitions (in the next to the month) and so on. Using these symbols, the probability distribution of the customer choosing any given and (D1, D2, D3) in any given month (k) may be expressed as a row vector as follows:

   Q(k) = [qD1(k) qD2(k) qD3(k)]

   In general, for n states

   Q(k) = [q1(k) q2(k) q3(k).........qn(k)]

   The initial condition is obviously expressed as Q(0).

   For the detergent example, since the market share for the three brands D1, D2, and D3 initially (on March is given to be 30%, 45% and 25%, respectively, we can write initial state probabilities as)

   Q(0) = [qD1(0) qD2(0) qD3(0) = [0.03 0.45 0.25]

   Now, the managers of the three brands of detergents would benefit from knowing the market shares that would our at a given future time (k). This information would be given by q(k) where k=1, 2, 3,........and so on. To be specific, let us calculate the share of the market likely to e held by each of the brands on April 1 (since the time period considered by us is one month, as mentioned earlier). This would e represented by Q(1), since k=1 for the next month. For this purpose, we use the matrix of transition probabilities, P. The row vector Q(0) would be post-multiplied by the matrix P to get Q(1). Thus,

0.60 0.30 0.10
Q(1)=(0.03 0.450.25) 0.20 0.50 0.30 = (0.3075 0.3275 0.3650)
0.15 0.05 0.80

To understand matrix multiplication, we may consider the first element 0.3075, which has been obtained as follows:

i. D1's share of the market x D1's propensity to retain its customers =0.30x0.60=0.1800

ii. D2's share of the market x D1's propensity to attract D2's customers = 0.45 x 0.20 =0.0900.

iii. D3's share of the market x D1's propensity to attract D3's customers= 0.25 x0.15= 0.0375

Total probability (i) +(ii)+ (iii) = 0.1800+0.0900+0

.0375 This is qD1(1).

In a similar way, the market share of 0.3275 for D2 is obtained as 0.30 x 0.30+0.45x 0.50+0.50+0.25 x0.05; while for D3, it is 0.30x0.10+0.45x0.30+0.25+0.80= 0.3650. These are, respectively, the values of qD2(1) and qD3(1).

In general terms, then,

Q(1) = [q1(1) q2(1) q3(1)............qn(1)] = Q(0)P

By similar reasoning, it may be shown that

Q(2) = Q(1)P, which may be expressed as

Q(2) = Q(0)PP, $Q(0)p^2$

In the same way, we cab obtainQ(3), Q(4) and so on. Thus,

Q(k) = Q(k-1) P=Q(k-2) p2 =..........=Q(0) P^K

Accordingly, the market shares of the three detergents two month hence, on May I, may be obtained as follows:

Thus market shares of the three brands of detergents D1, D2, and D3 are expected to be 30.5%, 27.49 and 42.1%, respectively, two months from now.

Besides calculating the market shares of the three brands at different points in time as represented by the state probabilities, information may also be obtained of the chances of a particular buyer buying a specific brand at a specific time, given that he previously purchased a particular brand. This situation is reflected when we desire to calculate, say , the probability that a customer, who has last purchased brand D2, would purchase brand D3 in two months from now. Such probabilities are termed as conditional state probabilities.

To calculate the probability of a customer to buy D3 two months hence, given that his latest purchase has been D2, we may recognize that the desired event of the buyer buying D3 may result in any of the following ways:

ii. Customer switches from brand D2 to D1 in month 1, and to brand D3 in month 2

iii. Customer stays on brand D2 in month 1, and switches to brand D3 in month 2

iv. Customer switches from brand D2 to D3 in month 1 and then stays on to D3 in month 2

The various probabilities are obtained from the transition probability matrix . For example, the probability of switching from D2 to D1 in a month is 0.20; D1 to D3 is 0.10; D2 toD3 is 0.30, and staying at D2 and D3 is 0.50 and 0.80, respectively. Since the transition probabilties are assumed to be constant over time, the probabilities for each month are obtained from the same matrix.

2. **Steady state probabilities:** A significant property of the Markov chains is that the process tends to stabilize in the long run. A stabilized system is said to be in a steady state or in equilibrium, so that the system's operating characteristics cab become independent of time. It is important to note that for a Markov chain to reach steady state condition, the chain must be ergodic. In an ergodic chain, it is possible to go from one state to any other state in a finite number of steps, irrespective of the present state. A transition probability matrix describing a Markov chain may be examined to see whether it is ergodic by checking if it is possible to move from every starting/ present state to all other states. In the brand switching example considered earlier, it is easy to see that one cab reach to every other state starting fro any state.

   The phenomenon of equilibrium cab be expressed symbolically as Q(k) = Q(k-1), so that the state probabilities in period k are the same as in the previous period. In the context of our detergent example, it means the following. We have seen in the preceding paragraphs as to how, given the current shares, the shares of various brands on month hence, two month hence, three month hence and so on can be calculated. We cab extend this process to any number of periods. However, it may e observed that as converge and stabilize. After this, successive calculations of qij for higher values of k do not show evidence of change. These are the equilibrium, or steady state probabilities.

Absorbing Chains

In our discussion so far, we have analyzed situations where none of the states was absorbing in nature. Now let us consider the Markov Chains where some of the states would be absorbing in nature. The accounts receivable situations of a firm may be modelled as an absorbing Markov chain.

For example: Consider the accounts receivable problem of the firm. Suppose that, for this firm, accounts receivable turns into bad debt if the account is more than three month overdue. Now, at the beginning of each month, each of the accounts may be classified into one of the following states:

State 1: New Account

State 2: Account is one month overdue for payment

State 3: Account is two month overdue for payment

State 4: Account is three month overdue for payment

State 5: Account has been paid

State 6: Account is written off as bad debt.

These ideas you will require in the coming lessons. You have to refer to the quantitative techniques and productions and operations management paper in your second and first semester for proper understanding.

MONTE CARLO TECHNIQUES OBJECTIVES

Upon completion of this lesson you will be able to understand:

- The meaning of simulation
- The phases of simulation
- The role and importance of monte-carlo simulation.

In the previous lesson, we have discussed mathematical models to describe and analyze the characteristics of a given system. Such models could be solved analytically to determine the optimal solutions or to describe the system. However, despite the fact that mathematical modeling can, and does, help to analyze a wide variety of problems, there are many situations which are too complex to be handled this way. For instance, not lend itself to modeling. In these circumstances, it is often possible to simulate the given system and study its behavior. Not only this, simulation may also be employed to estimate the impact of system changes when experimentation with the real system may not be feasible.

To simulate is to imitate. In general terms, simulation involves developing a model of some real phenomenon and then performing experiments on the model evolved. It is a descriptive, and not optimizing, technique. In simulation, a given system is copied and the variables and constants associated with it are manipulated in that artificial environment to examine the behavior of the system. To illustrate, for aerodynamic testing, scaled down models of aero planes are built and placed in the wind tunnels. Using a wind tunnel, air is blown to examine the aerodynamic properties of the model.

Using simulation, an analyst can introduce the constants and variables related to the problem, set up the possible courses of action and establish criteria which act as measures of effectiveness. The benefit of simulation from the viewpoint of the analyst stems from the fact that the results of taking a particular course of action cab be estimated prior to its implementation in the real world. Instead of using hunches and intuition to determine what may happen, the analyst using simulation cab test and evaluate various alternatives and select the one that gives the best result.

Process of Simulation

Broadly, there are four phases of the simulation process, they are

a. Definition of the problem and statement of objectives,

b. construction of an appropriate model

c. experimentation with the model constructed; and

d. Evaluation of the result of simulation.

Each of the phases called for the performance of a number of preliminary tasks. Of these, the two major tasks are collection of data and selection of means by which the simulation activity would replicate the random behavior of the real world.

The first step in problem solving of any situation is to edentify and clearly defines the problem and list the objectives (s) that the solution is intended to achieve. This is true of simulation as well. A clear statement not only facilitates the development of an appropriate model but also provides a basis for evaluation of the simulation results. In general, simulation aims to determine how the system under consideration would behave under certain conditions. Naturally, the more specific the analyst is about what he is looking for , the greater the chances that the simulation model will be designed to accomplish that. Thus, the scope and the level of detail of the simulation should be decided upon carefully.

The next step in simulation is the development of a suitable model. During the course of a simulation the model mimics the important elements of what is being simulated. A simulation model may be a physical or mathematical model, a mental conception, or a combination. Many simulations involve physical models. Examples include a scaled down model are relatively expensive to build, mathematical models are often preferred. In such a model, mathematical symbols or equations are used to represent the relationships in the system.

Collection of data is a significant aspect of model development, and the quantum and type of date needed are directly governed by the scope and extent of the details of the simulation. The data are needed both for model development and evaluation. Obviously, the model for simulation must be so designed experiments. This enables evaluation of the key decision alternatives. An ancillary step here is of designing experiments. The experiments help answer the 'what if...........' types of questions in simulation studies. By going through this process, the analyst is able to learn about the system behavior.

Once the simulation model is developed, the next step is to run it. If the model is model is deterministic, with all its parameters known and constant, then only a single run would suffice. On the other hand, if the simulation is stochastic in nature, with the parameters subject to random variation, then a number of runs would be needed to get a clear picture of the model performance. The probabilistic simulation is akin to the random sampling where each run represents on observation. Thus, statistical theory cab be used to determine the optimal sample sizes. Evidently, the greater the variability inherent in the simulation results, the larger would be the simulation runs needed to obtain a reasonable degree of confidence that the result are truly indicative of the system behavior.

The last step in the process of simulation is to analyze and interpret the result of the runs. The interpretation of results is , in a large measure, dependent on the extent to which the simulation model portrays the reality. Obviously, closer the approximation of the real system by the simulation model, lesser will be the need for adjusting the results and also lesser will be the risk inherent applying the result.

Monte Carlo Simulation

Although simulation can be of many types, our discussion will focus on the probabilistic simulation using the Monte Carlo method. Also called Computer simulation, it cab be described as a numerical technique that involves modeling a stochastic system with the objective of predicting the system's behavior The chance element is a very significant feature of Monte Carlo simulation and this approach cab be used when the given process has a random, or chance, component.

In using the Monte Carlo method, a given problem is solved by simulating the original date with random number generators. Basically, its use requires two things. First, as mentioned earlier, we must have a model that represents an image of the reality of the situation. Here the model refers to the probability distribution of the variable in question. What is significant here is that the variable may not be known to explicitly follow any of the theoretical distributions like Poisson, and so on. The distribution may be obtained by direct observation or from past records. To illustrate, suppose that a bakery keeps a record of the sale of the number of cakes of a certain type. Information relating to 200 days' sales is,

Demand (No. of cakes)

	:5	6	7	8	9	10	11	12	Total
(No of days)	4	10	16	50	62	38	12	8	200

Assuming that this is an adequate representation of the distribution of demand for the cake, we cab derive we probability distribution of demand by expressing each of the frequencies in terms of proportions. This is done by dividing each one of the values by 200- the total frequency. The resultant distribution follows:

Demand (No. of cakes):

	5	6	7	8	9	10	11	12Total	
(No of days)	0.20	0.50	0.80	0.25	0.31	0.19	0.60	0.04	

Thus, there is 0.02 or 2 per cent chance that 5 cakes would be demanded on a day, a 0.05 or 5 per cent hence that the demand would be for 6 cakes ... and son on. This distribution would serve as the model of the situation under consideration.

The second thing required for simulation as a mechanism to simulate the model-something to capture the random nature of the given system. Thus, we should have available a procedure that would help us to select, random, values for the variables which cab be used to approximate that state of the system. Such a mechanism cab be any random number generator consisting of a device or a procedure by which random members cab be determined and/or selected.

There are various ways in which random numbers (or apparently random, but not truly so) may be generated. These could be: result of some device like coin or die; published tables of random numbers, esquire method, or some other sophisticated method. It may be mentioned here that the 'random' numbers generated by some methods may not be really random in nature. In fact such numbers are called be shall not consider them be and consider only briefly how the numbers may be obtained and used.

One way to generate random numbers is to fix up a spinning arrow on a common clock. When the arrow spun, the number on which it stops would-be taken to be random number for that trial. Naturally, any number of spinning of the arrow would result in an equal number of random numbers. In s similar way, random numbers can be generated using spinning of a roulette wheel, tossing dice ......and like that. Although random numbers may be needed.

A more fast and convenient method is to make use of the published tables of random numbers, like the published by the Rand Corporation (of USA): a Million Random Digits. A random number table is a efficient way to generate random data in most situations. The numbers in this table are n random arguments. The underlying theory is that each number used for generating the random numbers. With computers, it is typically easier to generate random numbers is the didsquare. To illustrate this method, suppose that we wish to generate four-digit integers and the last number generated was 8937. To obtain the next number, in the quence, we square the las ton and use the middle four digits of the product. In this case the product is 869969 so that the next pseudo-number is 8699, having drawn up a suitable computer programme, a four-digit number may be fed into the computer and a list of pseudo-random numbers obtained.

Of all the random number generators, we shall make use of random number tables for demonstrating the solution process. In particular, we shall use the random number table (Table B7) given Appendix B of book. To consider how the table can be used for generating data relating to our bakery problem, we proceed as follows:

Step 1 An assignment has to be worked out so that the intervals of random numbers will correspond to the probability distribution. Here, since the probabilities have been calculated to two decimal places, which add up to 1.00, we need 100 numbers of two digits to represent each point of probability. Thus we take random numbers 00 through 99 to represent them. Now, as the probability of 5 cakes is equal to 0.02, we assign two random numbers 00-01 to this demand level; the probability of 6 cakes being equal to 0.05, the next five numbers, 02-06 would be assigned to this level. In a similar manner, each of the demand levels would be assigned appropriate intervals as given here. It may be mentioned that cumulative probabilities shown are calculated to ease the determination of the random number intervals. The cumulative probabilities column allows the assigned number to correspond to the same probability range for each event.

Demand (No. of cakes)	Probability	Cumulative Probability	Random No. Interval
5	0.02	0.02	00-01
6	0.05	0.07	02-06
7	0.08	0.15	07-14
8	0.25	0.40	15-39
9	0.31	0.71	40-70
10	0.19	0.90	71-89
11	0.06	0.96	90-95
12	0.04	1.00	96-99

Instead, probabilities are calculated to three decimal places, then 3- digit random numbers would be required ......and son on.

Step 2 Once the random number intervals are determined, we select a tracking pattern for drawn random numbers from the random number table. We may start with any column and row of the table and read the values in any set manner – horizontally, vertically, or diagonally. Using the pattern, we draw the random numbers and match them with the assigned events. We may decide, for example, to read every third value horizontally, starting with the fifth column and fourth row of the table of random numbers. The random numbers, according to this pattern are 60,74,24.03,59,16,84,92,52,07 and so on. We draw as many random numbers as the number of days' demand is required to be simulated.

The first of the list of the numbers, 61, lies in the interval 40-70 corresponding to the demand level 9 units. Thus, the simulated demand for the first day is 9 cakes. In a similar manner, we can obtain the demand for each of the days. For the 10-day period, we have the following demand:

Day :	1	2	3	4	5	6	7	8	9	10
Random Number:	61	74	24	03	59	16	84	93	52	07
Demand (cakes):	9	10	8	6	9	8	10	11	9	7

These are the only parts which we need to understand in the area of financial engineering. For your purpose you need to study it deeply because it is going to use these concepts in the coming analysis.

Okey.

TUTORIAL

OVERVIEW OF FINANCIAL MARKET

Objective

Upon completion of this lesson you will be able to understand different financial markets throughout the world.

Hello, today we will come to the necessary groundwork for studying the methods of financial engineering. Readers with a good grasp of the conventions and mechanics of financial markets may skip it, although a quick reading would be preferable.

You have seen that financial engineering is a practice and can be used only when we define the related environment carefully. The organization of markets, and the way deals are concluded and earned out, are important factors in selecting the right solution for a particular financial engineering problem. This lesson examines the organization of financial markets and the way market practitioners interact. Issues related to settlement, to accounting methods, and especially to conventions used by market practitioners are important and need to be discussed carefully.

In fact, it is often overlooked that financial practices will depend on the conventions adopted by a particular market. This aspect, which is relegated to the background in most books, will be an important parameter of our approach. Conventions are not only important in their own right for proper pricing, but they also often reside behind the correct choice of theoretical models for analyzing pricing and risk management problems. The way information is provided by markets determines the model choice, and this cannot be discussed before briefly reviewing some basic conventions. While doing this, the chapter introduces the mechanics of the markets, instruments, and who the players are. A brief discussion of the syndication process is also provided.

Markets

The first distinction is between local and Euromarkets. Local markets are also called onshore markets. These denote markets that are closely supervised by regulators such as central banks and financial regulatory agencies. There are basically two defining characteristics of onshore markets. The first is reserve requirements that are imposed on onshore deposits. The second is the formal registration process of newly issued securities. Both of these have important cost, liquidity, and taxation implications. Reserve requirements imposed on banks increase the cost of holding onshore deposits and making loans. This is especially true of the large "wholesale" deposits that banks and other corporations may use far short periods of time, Short maturities normally command low interest. If part .of these funds are held in a non interest-bearing farm in central banks, the cost of local funds will increase.

The long and detailed registration process imposed an institution that is issuing stacks, bonds, or other financial securities have two implications for financial engineering. First, issue casts will be higher in case of registered securities when compared to simpler bearer form securities. Second, an issue that dues nut have to be registered with a public entity will disclose less information.

Thus, markets where reserve requirements do not exist, where the registration process is simpler, will have significant cast advantages. Such markets are called Euromarkets.

Euromarkets

We should set something clear at the .outset. The term "Euro" as utilized in this section does not refer to Europe, nor does it refer to the Euro zone currency, the Euro. It simply means that, in terms .of reserve requirements or registration process we are dealing with markets that are outside the formal control of regulators and central hanks. The two most important Euromarkets arc the Eurocurrency market and the Eurobond market.

Eurocurrency Markets

Start with an onshore market. In an onshore system, a 3-m.onth retail deposit has the following life. A client will deposit USD 100 cash on date T. This will be available the same day. That is to say, "days to deposit" will equal zero. The deposit-receiving bank takes the cash and deposits, say, 10% of this, in the central bank. This will be the required reserves portion of the original 100. The remaining 90 dollars are then used to make new loans or may be lent to other banks in the inter-bank overnight market. Hence, the bank will be paying interest on the entire 100, but will be receiving interest on only 90 of the original deposit. In such an environment, assuming there is no other cost, the bank has to charge an interest rate around ten percent higher far making loans. Such supplementary costs are enough to hinder liquid wholesale market far money where large sums are moved. Eurocurrency markets eliminate these casts and increase the liquidity.

Let's briefly review the life .of a Eurocurrency (offshore) deposit and compare it with an onshore deposit. Suppose a U.S. bank deposits USD 100 million in another U.S. bank in the New York Eurodollar (offshore) market. Thus, as is the case for Eurocurrency markets, we are dealing .only with banks, since this is an inter-bank market. Also, in this example, all banks are located in the United States. The Euro deposit is made in the United States and the "money" never leaves the United States. This deposit becomes usable (settles) in 2 days-that is to say, days to deposit are 2 days. The entire USD 100 million can now be lent to another institution as a loan. If this chain of transactions was happening in, say, London, the steps would be similar.

Eurobond Markets

A bond sold publicly by going through the formal registration process will be an onshore instrument. If the same instrument is sold without a similar registration process, say, in London, and if it is a bearer security, then it becomes essentially an offshore instrument. It is called a Eurobond. Again the prefix "Euro" does not refer to Europe, although in this case the center of , Eurobond activity happens to be in London. But in principle, a Eurobond can be issued in Asia as well.

A Eurobond will be subject to less regulatory scrutiny, will be a bearer security, and will not be (as of now) subject to withholding taxes, The primary market will be in London, The secondary markets may be in Brussels. Luxembourg or other places where the Eurobonds will be listed. The settlement of Eurobonds will be done through Euro-clear or cedel.

Other Euromarkets

Euromarkets are by no means limited to bonds and currencies, almost any instrument can be marketed offshore. There can be Euro-equity, Euro-commercial paper (ECP), Euro medium-term note (EMTN), and so on.

Onshore Markets

Onshore markets can be organized over the counter or as formal exchanges, Over-the-counter (OTC) markets have evolved as a result of spontaneous trading activity, An OTC market often has no formal organization, although it will be closely monitored by regulatory agencies and transactions may be carried out along some precise documentation drawn by, professional organizations, such as ISDA, ISMA.[3] Some of the biggest markets in the world are OTC. A good example is the interest rate swap (IRS) market, which has the highest notional amount traded among all financial markets with very tight bid-ask spreads. OTC transactions are 'often done over the phone and the instruments contain a great deal of flexibility, although, again, institutions such as International Swaps and Derivatives Association draw standardized documents that make traded instruments homogeneous.

In contrast to OTC markets, organized exchanges are formal entities; they may be electronic or open-outcry exchanges. The distinguishing characteristic of an organized exchange is its formal organization. The traded products and trading practices are homogenous while, at the same time, the specifications of the traded contracts are less flexible.

A typical deal that goes through a traditional open-outcry exchange can be summarized as follows:

1. A client uses a standard telephone to call a broker to place an order, The broker will take the order down.
2. Next, the order needs to be transmitted to exchange floors or, more precisely, to a booth.
3. Once there, the order needs to be sent out to the pit, where the actual trading is done.
4. Once the order is executed in the pit a verbal confirmation process needs to he implemented all the way hack to the client.

Stock markets are organized exchanges that deal in equities. Futures and options markets process derivatives written on various underlying assets. In a spot deal, the trade will be done and confirmed, and within a few days, called the settlement period, money and securities change hands. In futures markets, on the other hand, the trade will consist of taking positions, and settlement will be after a relatively longer period, once the derivatives expire. The trade is, however, followed by depositing a "small" guarantee, called an initial margin.

Different exchanges have different structures and use different approaches in Market Making. For example, at the New York Stock Exchange (NYSE), market making is based on the specialist system. Specialists run books on stocks that they specialize in. As market makers, specialists are committed to buying and selling at all times at the quoted prices and have the primary responsibility of guaranteeing a smooth market.

Futures Exchanges

EUREX, CBOT, CME, LIFFE, and TIFFE are some of the major futures and options exchanges in the world. The exchange provides three important services:

1. A physical location (i.e., the trading floor and the accompanying pits) for such activity, if it is an open-outcry system. Otherwise the exchange will supply an electronic trading platform.
2. An exchange clearinghouse that becomes the real counterparty to each buyer and seller once the trade is done and the deal ticket is stamped.
3. The service of creating and designing financial contracts that the trading community needs and, finally, providing a transparent and reliable trading environment.

The mechanics of trading in futures (options) exchanges is as follows. Two pit traders trade directly with each other according to their client's wishes. One sells, say, at 100; the other buys at 100. Then the deal ticket is signed and stamped. Until that moment, the two traders are each other's counterparties. But once the deal ticket is stamped, the clearinghouse takes over as the counterparty. For example if a client has bought a futures contract for the delivery of 100 bushels of wheat, then the entity responsible for delivering the wheat is not the "other side" who physically sold the contract on the pit, but the exchange clearinghouse. By being the only counterparty to all short and long positions, the clearinghouse will lower the counterparty risk dramatically. The counterparty risk is actually reduced further, since the clearinghouse will deal with clearing members rather than the trader directly.

An important concept that needs to be reviewed concerning futures markets is the process of marking to market. When one "buys" a futures contract, a margin is put aside, but no cash payment is made. This leverage greatly increases the liquidity in futures markets, but it is also risky. To make sure that counterparties realize their gains and losses daily, the exchange will reevaluate positions every day using the settlement price observed at the end of the trading days.

Example

A 3-month Eurodollar futures contract has a price of 98.75 on day T. At the end of day T + 1, the settlement price is announced as 98.10. The price fell by 0.65, and this is a loss to the long position holder. The position will be marked to market, and the clearinghouse-or more correctly-the clearing firm, will lower the client's balance by the corresponding amount.

The open interest in futures exchanges is the number of outstanding futures contracts. It is obtained by totaling the number of short and long positions that have not yet been closed out by delivery, cash settlement, or offsetting long/short positions.

MARKET PLAYERS AND CONVENTIONS

Objectives

On completion of this lesson you would be able to understand the following objectives

- Understanding of the market players
- To identify the conventions and structure
- Understanding of the instruments and characteristics

Players

Market makers warehouse financial instruments and provide the traders with two-way quotes. They provide liquidity and smooth out severe market fluctuations. Market makers must as an obligation, buy and sell at their quoted prices. Thus, for every security at which they are making the market, the market maker must quote a bid and an ask price. A market maker does not warehouse a large number of products, nor does the market maker hold them for a long period of time.

Traders buy and sell securities. They do not, in the pure sense of the word, "make" the markets. A trader's role is to execute clients' orders and trade for the company given his or her position limits. Position limits can be imposed on the total capital the trader is allowed to trade' or on the risks that he or she wishes to take.

A trader or market maker may run a portfolio, called a book. There are "FX books," "options books," "swap books," and "derivatives books," among others. Books run by traders are called "trading books"; they are different from "investment portfolios," which are held for the purpose of investment. A trading book carries instruments because during the process of buying and selling for clients, the trader may have to warehouse these products for a short period of time. These books are hedged periodically.

Brokers do not hold inventories. Instead, they provide a platform where the buyers and sellers' can get together. Buying and selling through brokers is often more discreet than going to bids and asks of traders. In the latter case, the trader would naturally learn the identity of the client.

In options markets, a floor-broker is a trader who takes care of a client's order but does not trade for himself or herself. (On the other hand, a market maker does.)

Dealers quote two-way prices and hold large inventories of a particular instrument maybe, for a longer period of time than a market maker. They are institutions that act in some sense as market makers.

Risk managers are relatively new players. Trades, and positions taken by traders, should be "approved" by risk managers. The risk manager assesses the trade and gives approvals if the trade remains within the pre-selected boundaries on various risks.

Regulators are important players in financial markets. Practitioners often take positions of"tax arbitrage" and "regulatory arbitrage." A large portion of financial engineering practices are directed toward meeting the needs of the practitioners in terms of regulation and taxation.

Researchers and analysts are players who do not trade or make the market. They are information providers for the institutions and are helpful in sell-side activity. Analysts in general deal with stocks and analyze one or more companies. They can issue buy/ sell/hold signals and provide forecasts. Researchers provide macro level forecasting and advice.

The Mechanics of Deals

What are the mechanisms by which the deals are made? How are trades done? It turns out that organized exchanges have their own clearinghouses and their own clearing agents. So it is relatively easy to see how accounts are opened, how payments are made, how contracts are purchased and positions are maintained. The clearing members and the clearinghouse do most of these. But how are these operations completed in the case of OTC deals? How does one buy a bond and pay for it? How does one buy a foreign currency?

Turning to another detail, where are these assets to be kept? An organized exchange will keep positions for the members, but who will be the custodian for GTC operations and secondary market deals in bonds and other relevant assets? Several alternative mechanisms are in place to settle trades and keep the assets in custody. A typical mechanism is shown in the following figure.

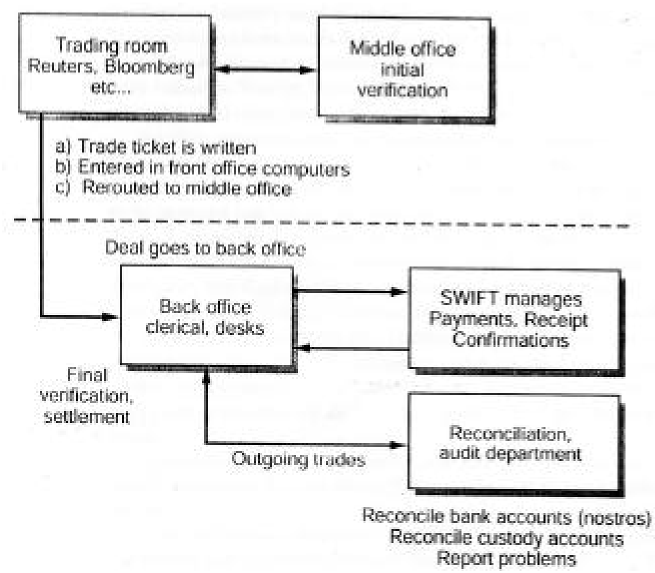
The mechanics of a deal in this figure are from the point of view of a market practitioner. The deal is initiated at the trading or dealing room. The trader writes the deal ticket and enters this information in the computer's front office system. The middle office is the part of the institution that initially verifies the deal. It is normally situated on the same floor as the trading room. Next, the deal goes to the back office, which is located either in a different building or on a different floor. Back-office activity is as important for the bank as the trading room. The back office does the final verification of the deal handles settlement instructions, releases payments, and checks the incoming cash flows, among other things. The back office will also handle the messaging activity using the SWIFT system, to be discussed later.

Orders

There are two general types of orders investors or traders can place. The first is a market order, where the client gets the price market quotes at that instant.

Alternatively one can place a *limit order*. Here a price will be specified along the order, and the trade will go through only if this or a better price is obtained. A limit order is valid only during a certain period, which needs to be specified also. A *stop loss* order is similar. It specifies a target price at which a position gels liquidated automatically.

Processing orders is by no means error-free. For example, one disadvantage of traditional open-outcry exchanges is that in such an environment, mistakes are easily made, Buyer and



seller may record different prices. This is called a “price out” Or there may be a “quantity out,” where the buyer has “bought 100” while the seller thinks that he has “sold 50.” In the case of options exchanges, the recorded expiration dates may not match, which is called a “time out” *Out-trades* need to be corrected after the market close. There can also be *missing trades.* These trades need to be negotiated in order to recover positions from counterparties’ and clients.

4.2. Confirmation and Settlement

Order confirmation and settlement are two integral parts of financial markets. Order confirmation involves sending messages between counterparties, to confirm trades verbally agreed upon between market practitioners. Settlement is exchanging the cash and the related security, or just exchanging securities.

The SWIFT system is a communication network that has been created for “paperless” communication between market participants. It stands for the Society for Worldwide Financial Telecommunications and is owned by a group of international banks. The advantage of SWIFT is the standardization of messages concerning various transactions such as customer transfers, bank transfers, Foreign Exchange (FX), loans, deposits. Thousands of financial institutions in more than 100 countries use this messaging system.

Another interesting issue is the relationship between settlement, clearing, and custody. Settlement means receiving the security and making the payment the institutions can settle, but in order for the deal to be complete, it must be cleared. The orders of the two counterparties need to be matched and the deal terminated. Custody is the safekeeping of securities by depositing them with carefully selected depositories around the world. A custodian is an institution that provides custody services. Clearing and custody are both rather complicated tasks. FedWire, Euroclear, and Cedel are three international securities clearing firms that also provide some custody services. Some of the most important custodians are banks.

Countries also have their own clearing systems. The best known clearing systems are CHIPS and CHAPS. CHAPS is the clearing system for the United Kingdom, CHIPS is the clearing system for payments in the United States. Payments in these systems are cleared multilaterally and payments are netted. This greatly simplifies settling large numbers of individual trades.

Spot trades settle according to the principle of DVP - that is to say, delivery versus payment - which means that first the security is delivered (to securities clearing firms) and then the cash is paid.

Issues related to settlement have another dimension. There arc important conventions involving normal ways of settling deals in various markets. When a settlement is done according to the convention in that particular market, we say that the trade settles in a regular way. Of course, a trade can settle in a special way. But special methods would be costly and impractical.

Example

Market practitioners denote the trade date by T. and settlement is selected relative to this date. U.S. Treasury securities settle regularly on the first business day after the trade-that is to say, on T + 1. But it is also common for efficient clearing firms to have cash settlement-that is to say settlement is done on the trade date T.

Corporate bonds and international bonds settle on T + 3.

Commercial paper settles the same day.

Spot transactions in stocks settle regularly on T + 3 in the United States.

Euro-market deposits are subject to T + 2 settlements. In the case of overnight borrowing and lending, counterparties may choose cash settlement.

Foreign exchange markets settle regularly on T + 2. This means that a spot sale (purchase) of a foreign currency will lead to two-way flows two days after the trade date, regularly. T + 2 is usually called the spot date.

Another issue must be remembered in these settlement conventions. The number of days to settlement in general refers to business days. This means that in order to be able to interpret T + 2 correctly, the market professional would need to pin down the corresponding holiday convention.

Before discussing other market conventions, we can mention two additional terms that are related to the preceding dates. The settlement date is sometimes called the value date in contracts. Cash changes hands at the value date. Finally, in swap-type contracts, there will be the deal date (Le. when the contract is signed), but the swap may not begin until the effective date. The latter is the actual start date for the swap contract and will be at an agreed-upon later date.

Market Conventions

Market conventions often cause confusion in the study of financial engineering. Yet, it is very important to be aware of the conventions underlying the trades and the instruments. In this section, we briefly review some of these conventions.

Conventions vary according to the location and the type of instrument one is concerned with. Two instruments that are quite similar may be quoted in very different ways. *What* is quoted and *the way* it is quoted are important.

As mentioned, in Chapter I in financial markets there arc always *two* prices. There is the price at which a market maker is willing to *buy* the underlying asset and the price at which he or she is willing to *sell* it. The price at which the market maker is willing to buy is called the *bid* price. The *ask* price is the price at which the market maker is willing to sell. In London financial markets, the ask price is called an *offer*. Thus, the bid-ask spread becomes the bid-offer spread. As an example consider the case of deposits in London *money and foreign exchange markets,* where the convention is to quote the *asking* interest Tate first. For example, a typical quote on interest rates would be as follows:

Ask (offer)	Bid
5 ¼	5 1/8

In other money centers, interest rates are quoted the other way around. The first rate is the bid, the second is the ask rate. Hence, the same rates will look as such:

Ask (offer)	Bid
5 1/8	5 ¼

A second characteristic of the quotes is decimalization. The Eurodollar interest rates in London are quoted to the nearest 1/16 or sometimes 1/32. But many money centers quote interest rates to *two* decimal points. Decimalization is not a completely straightforward issue from the point of view of brokers/dealers. Note that with decimalization, the bid-ask spreads may narrow all the way down to zero, and there will be no minimum bid-ask spread. This may mean lower trading profits, everything else being the same. Also, the reader must be aware that the decimalization characteristics may change over time.

What to Quote

Another set of conventions concerns what to quote. For example, when a trader receives a call he or she might say, "I sell a bond with price 95," or instead, he or she might say, "I sell a bond with yield 5%." Markets prefer to work with conventions to avoid potential misunderstandings and to economize time. Equity markets quote individual stock prices. On the New York Stock Exchange the quotes are to decimal points.

Most bond markets quote prices rather than yields, with the exception of short-term T-bills. For example, the price of a bond may be quoted as follows:

Bid price	Ask (Other) price
90.45	90.57

The first quote is the price a market-maker is willing to pay for a bond. The second is at which the market-maker dealer is willing to sell the same bond. Note that according to this, bond prices are quoted to *two* decimal points, out of a par value of 100, regardless of the true denomination of the bond.

It is also possible that a market quotes neither a price nor a yield. For example, caps, floors, and swaptions often quote "volatility" directly. Swap markets prefer to quote the "spread" (in the case of USD swaps) or the swap rate itself (Euro-denominated swaps). The choice of what to quote is not a trivial matter. It affects pricing as well as risk management.

How to Quote Yields

Markets use three different ways to quote yields. These are, respectively, the money market yield, the bond equivalent yield, and the discount rate[7]. We will discuss these using default-free "pure discount bonds with maturity T as an example. Let the time-t price of this bond be denoted by B (t, T). The bond is default free and pays 100 at time T. Now, suppose R^T represents the time-t yield of this bond.

It is clear that B(t,T) will be equal to the present value of 100, discounted using R^T, but how should this present value be expressed? For example, assuming that (T - t) is measured in days and that this period is less than 1 year we can use the following definition:

$$B(t,T) = \frac{100}{(1+R^T)^{(T-t/365)}}$$

where the $^{(T-t/365)}$ is the remaining life of the bond as a fraction of year, which, here is "defined" as 365 days.

This latter term is different from the special interest rate used by the U.S. Federal Reserve System, which carries the same name. Here the discount rate is used as a general category of yields. But we can also think of discounting the maturity value using the alternative:

$$B(t,T) = \frac{100}{(1+R^T)^{(T-t/365)}}$$

Again, suppose we use neither formula but instead set

$$B(t,T) = 100 - R^T \frac{T-t}{365}$$

Some readers may think that given these formulas, (1) is the right one to use. In fact, they may *all* be correct, given the proper convention.

The best way to see this is to consider a simple example. Suppose a market quotes prices *B (t, T)* instead of the yields *RT*. _ Also suppose the observed market price is

$$B(t, T) = 95.00$$

with *(T* - t) = 180 days and the year having 365 days. We can then ask the following question: Which one of the formulas in (1) through (3) will be more correct to use? It turns out that these formulas can *all* yield the same price 95.00, *if* we allow for the use of different yields.

In fact, with R^T = 10.9613% the first formula is "correct," since

$$B(t,T) = \frac{100}{(1+.109613)^{(180/365)}}$$

$$= 95.00$$

On the other hand, with R^T_2= 10.6725% the second formula is "correct," since

$$B(t,T) = \frac{100}{(1+.106725)^{(180/365)}}$$

$= 95.00$

Finally, if we let R^T_3= 10.1389% the third formula is "correct," since

$$B(t,T) = 100 - 101389^{(180/365)} 100$$

$$= 95.00$$

Thus, for (slightly) different values of R^T, all formulas give the same price. But which one of these is the right formulas give the same price. But which one of these is the "right" formula to use?

That is exactly where the notion of convention comes in. A market can adopt a convention to quote yields in terms of Formula (1). Then, once traders see a quoted yield in this market, they would "know" that the yield is defined in terms of Formula (1) and not by (2) or (3). This convention, which is only an implicit understanding during the execution of trades, will be expressed precisely in the actual contract and will be known by all traders. A newcomer to a market can make serious errors if he or she does not pay enough attention to market conventions.

Emerging market bonds are in general quotes in terms of yields. In treasury markets, the quotes are in terms of prices. This may make some difference from the point of view of both market psychology pricing and risk management decisions.

Example

In the United States bond markets quote the yields in terms of Formula (I). Such values of R^{T1} are called bond equivalent yields.

Money markets that deal with inter-bank deposits and loans use the money-market yield convention and utilize Formula (2) in pricing and risk management.

Finally, the Commercial Paper and Treasury Bills yields are quoted in terms of Formula (3). Such yields are called discount rates.

Finally, the continuous discounting and the continuously compounded yield r is defined by the formula

$$B(t.\ T) = 100c^{r}(T - t)$$

where the e^x is the exponential function. It turns out that markets do not like to quote continuously compounded yields. One exception is toward retail customers. Some retail bank accounts quote the continuously compounded savings rate. On the other hand, the continuously compounded rate is often used in some theoretical models and was until lately, the preferred concept for academics. .

One final convention needs to be added at this point. Markets have an interest payments convention as well. For example, the offer side interest rate on major Euroloans, the Libor, is paid at the conclusion of the term of the loan as a single payment. On the other hand, many bonds make periodic coupon payments that occur on dates earlier than the maturity of the relevant instrument.

Day, Count Conventions

The previous discussion suggests that ignoring conventions can lead to costly numerical errors in pricing and risk management. A similar comment can be made about *day count* conventions. A financial engineer should immediately check the relevant day count rules in the products that he or she is working on. The reason is simple. The definition of a "year" or of a "month" may change from one market to another and the quotes that one observes will depend on this convention. The major day-count conventions are as follows:

1. The 30/360 basis. Every month has 30 days regardless of the actual number of days in that particular month, and a year has always 360 days For example, an instrument following this convention and purchased on May I and sold on July 13 would earn interest on

   30 + 30 + 12 = 72 days,

   while the actual calendar would give 73 days.

   More interestingly, this instrument purchased on February 28, 2003, and sold the next day, on March 1, 2003, would earn interest for 3 days. Yet, a money market instrument such as an inter-bank deposit would have earned interest on only I day (using the aetual/360 basis mentioned below).
2. The 30E/360 basis. This is similar to 30/360 except for a small difference at the end of the month, and it is used mainly in the Eurobond markets. The difference between 30/360 and 30E/360 is illustrated by the following table, which shows the number of ways interest is earned starting from March I according to the two conventions:

Convention	March I – March 30	March I – March 30	March I-April I
30E/360	29 days	29 days	30 days
30/360	29 days	30 days	30 days

   According to this, a Eurobond purchased on March I and sold on March 31 gives an extra day of interest in the case of 30/360, whereas in the case of 30E/360, one needs to hold it until the beginning of the next month to get that extra interest.
3. The actual /360 basis. If an instrument is purchased on May 1 and sold on July 13, then it is held

   73 days under this convention. This convention is used by most money markets.
4. The actual / 360 basis. This is the case for Euro sterling money markets, for example.
5. Actual/actual. Many bond markets use this convention.

An example will show why these day count conventions are relevant for pricing and risk management. Suppose you are involved in an interest rate swap. You pay Libor and receive fixed. The market quotes the Libor at 5.01, and quotes the swap rate at 6.23/6.27. Since you are receiving fixed, the relevant cash flows will come from paying 5.01 and receiving 6.23 at regular intervals. But these numbers are somewhat misleading. It turns out that Libor is quoted on an *ACT* /360 basis. That is to say, the number 5.01 assumes that there are 360 days in a year. However, the swap rates may be quoted on an *ACT* /365 basis, and all calculations may be based on a 365-day year[9]. Also the swap rate may be *annual* or *semiannual.* Thus, the two interest rates where one pays 5.01 and receives 6.23 are *not* directly comparable.

Example

Swap markets are the largest among all financial markets, and the swap curve has become the central pricing and risk management roll in finance Hence. It is worth discussing swap market conventions briefly.

- USD swaps are liquid against 311l-Libor and 6m-Libor. The day-count basis is annual, ACT/360.
- Japanese Yen (JYP) swaps are liquid against 6m-Libor. The day-count basis is semiannual, ACT/365.
- British pound (GBP) swaps are semiannual, ACT/365 versus 6m-Libor.
- Finally, Euro (EUR) swaps are liquid against 6m-Libor and against 6m-Euribor.The day-count basis is annual 30/360.

Table above summarizes the day count and yield/discount conventions for some important markets around the world.

A few comments are in order concerning this table. First note that the table is a summary of three types of conventions. The first is the day-count, and this is often ACT/360. However, when the 30/360 convention is used, the 30E/360 version is more common. Second, the table tells us about the yield quotation convention. Third, we also have a list of coupon payment conventions concerning long-term bonds. Often these involve semiannual coupon payments.[10]

Finally, note that the table also provides a list of the major instruments used in financial markets. The exact definitions of these will be given gradually in the following chapters.

5.3.1. Holiday Conventions

Financial trading occurs across borders. But holidays adopted by various countries are always somewhat different. There are special independence days, special religious holidays. Often during Christmas time, different countries adopt different holiday schedules. In writing financial contracts, this simple point should also be taken into account, since we may not receive the cash we were counting on if our counterparty's markets were closed due to a special holiday in that country.

Hence, all financial contracts stipulate the particular holiday schedule to be used (London, New York, and so on), and then specify the date of the cash settlement if it falls on a holiday. This could be the next business day or the previous business day, or other arrangements could be made.

Two Examples

We consider how day-count conventions are used in two important cases. The first example summarizes the confirmation of short term money market instruments, namely a Eurodollar deposit. The second example discussed the confirmation summary of a Eurobond.

EXAMPLE	A Eurodollar Deposit
Amount	$ 100,000
Trade date	Tuesday, June 5, 2002
Settlement date	Thursday, June 7, 2002
Maturity	Friday, July 5, 2002
Days	30
Offer Rate	4.789%

Interest concerned (100,000) X 0.04789 X 30/60

Note three important points. First, the depositor earns interest on the settlement date, but does not earn interest for the day contract matures. This gives 30 days until maturity. Second, we are looking at the deal from the bank's side, where the bank sells a deposit, since the interest rate is the offer rate. Third, note that interest is calculated using the formula

$$(1 + r\delta)\,100{,}000 - 100{,}000$$

and not according to

$$(1 + r)^{\delta}\quad 100{,}000 - 100{,}000$$

where = 30/360
is the day-count adjustment.

The second example involves a Eurobond trade.

Example. A Eurobond

April 25, 2008

European Investment Bank, 5.0% (Annual Coupon)

Trade date	Tuesday, June 5, 2002
Settlement date	Monday, June 11, 2002
Maturity	December, 28, 2006
Previous coupon	April 25, 2001
Next coupon	April 25, 2002
Days in coupon period	360
Accrued coupon	Calculate using money market yield

We have two comments concerning this example. The instrument is a Eurobond, and Eurobonds make coupon payments annually, rather than semi-annually (as in the case of Treasuries, for example). Second, the Eurobond year is 360 Jays. Finally, accrued interest is calculated the same way as in money markets.

6. Instruments

   This section provides a list of the major instrument classes from the perspective of financial engineering. A course on markets and instruments along the lines of Hull (2002) is needed for a reasonable understanding.

The convention in financial markets is to divide these instruments according to the following sectors:

1. Fixed income instruments. These are interbank certificate of deposits (CDs), or deposits (depos), commercial paper (CP), banker's acceptances, and Treasury bills. These _re considered to be

   money market instruments.

   Bonds, notes, and Floating Rate Notes (FRNs) are bond market instruments.
2. Equities. These are various types of stock issued by public companies.
3. Currencies and commodities.
4. Derivatives, the major classes of which are interest rate, equity. currency. and commodity derivatives.
5. Credit instruments, which arc mainly high-yield bonds, corporate bonds, credit derivatives, and various guarantees that are early versions of the former.

We discuss these major classes of instruments from many angles in the lessons that follow.

Positions

By buying or short-selling assets, one takes *positions,* and once a position is taken, one has *exposure* to various risks.

Short and Long Positions

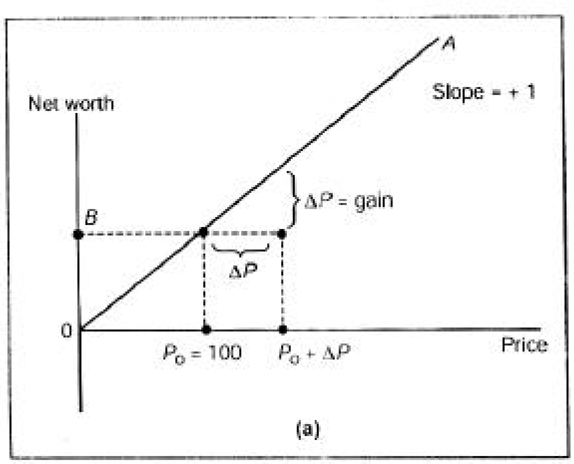
A *long* position is easier to understand, because it conforms to the instincts of a newcomer to financial engineering. In their daily lives, individuals often "buy" things; they rarely "short" them. Hence, when we buy an item for cash and hold it in inventory, or when we sign a contract that obliges us to buy something at a future date, we will have a long position. We are long the "under lying instrument," and this means that we benefit when the value of the underlying asset increases.

A *short* position, on the other hand, is one where the market practitioner has sold an item without really owning it. .For example, a client calls a bank and buys a particular bond. The bank may not have this particular bond on its books, but can still *sell* it. In the meantime, however, the hank has a short position.

A short position can be on an *instrument,* such as selling a "borrowed" bond, a stock, a future commitment, a swap, or an option. But the short position can also be on a particular *risk*. For example, one can be short (long) volatility - a position such that if volatility goes up, we lose (gain). Or one can be short (long) a spread - again, a position where if the spread goes up, we lose (gain).

Payoff Diagrams

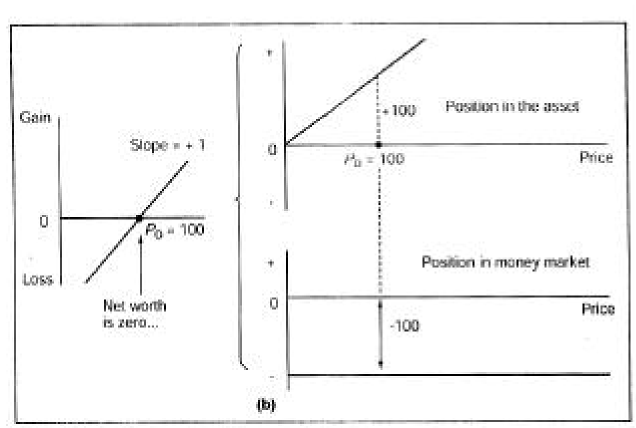
One can represent short and long positions using *payoff diagrams.* The following figure illustrates the long position from the point of view of an investor. The upward-sloping vertical line *OA* represents the value of the investor's position given the price of the security. Since its slope is + 1, the price of the security *P* will also be the value of the initial position. More correctly, note that if starting from P_0 the price increases by D*P*, the gain will be equal to this change as well.



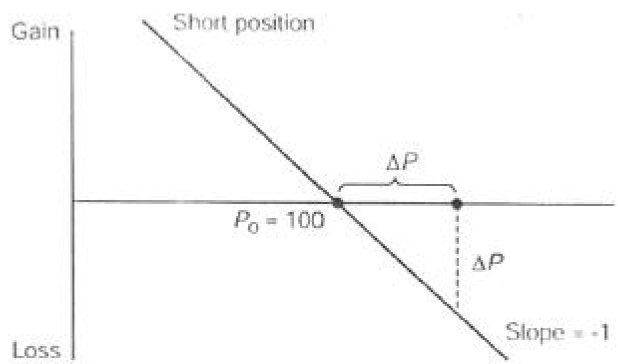
(a)

In particular, if the investor "buys" the asset when the price is 100 using his or her own savings, the net worth at that instant is represented by the vertical distance OB, which equals 100. A market professional, on the other hand, has no "money". So he or she has to borrow the OB (or the P_0) first and then buy the asset. This is called funding the long position.

This situation is shown in Figure. Note that the market professional's total net position amounts to zero net worth at the time of the purchase when the price is 100. In a sense, by first borrowing and then buying the asset, one "owns" not the asset but some *exposure.* If the asset price goes up, the position becomes profitable. If, on the other hand, the price declines, the position will show a loss.



(b)



The figure above shows a short position from a market practitioner's point of view. Here the situation is simpler. The asset in the short position is borrowed anyway. Hence, when the price is 100 at the time of the sale, the net worth is automatically zero. What was sold was an asset that was worth 100. The cash generated by the sale just equals the value of the asset that was borrowed. Therefore, at the price *P* = 100, the position has zero value. The position will gain when the price *falls* and will lose when the prices goes *up.* This is the case since what is borrowed is a security and not "money." Furthermore, this asset is sold at 100. Hence, when the asset price increases, one would have to return to the original owner a security that is worth *more* than 100.

Similarly, when the security price falls, one *covers* the short position by buying a new security at a price lower than 100 and then returning this (less valuable) asset to the original owner.

Overall, the short position is described by a downward sloping straight line with slope – 1.

It is interesting to note some technical aspects of these graphs. First, the payoff diagrams that indicate the value of the positions taken are linear in the price of the asset. As the price P changes, the payoff changes by a constant amount. This sensitivity of the position to price changes is called delta. In fact, given that the change in price will determine the gains or losses" on a one-to-one basis, the delta of a long position will be 1. In the case of a short position; the delta will equal -1.

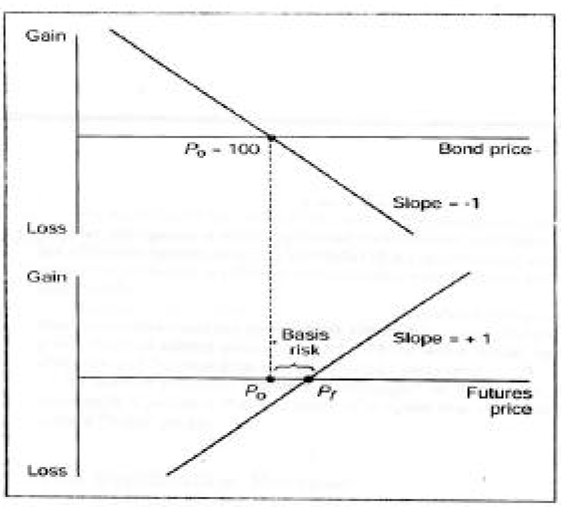
One can define many other sensitivity factors by taking other partial derivatives. Such sensitivities are called Greeks and are extensively used in option markets.[11]

Types of Positions

Positions can be taken for the purposes of hedging, arbitrage, and speculation. We briefly review these activities.

Let us begin with hedging. Hedging is the act of eliminating the exposures of existing positions without unwinding the position itself. Suppose we are short a bond (i.e., we borrowed somebody's bond and sold it in the market for cash). We have cash at hand, but at the same time, we owe somebody a bond. This means that if the bond price goes up, our position will have a mark-to-market loss.

11 Note that bid-ask spreads are not factored in the previous diagrams. The settling and buying prices cannot be the same at 100. The selling price P^{ask} will he larger than the buying price P^{bid} the P^{ask} - P^{bid} will be the corresponding bid-ask spread.



In order to eliminate the risk we can buy a "similar" bond. Our final position is shown in Figure. The long and short positions "cancel" each other except for some remaining basis risk. At the end, we will have little exposure to movements in the underlying price *P.* To hedge the same risk we can also take the long position not in the *cash* or *spot* bond markets, but in a *futures* or *forward* market. That is to say, instead of buying another bond, we may write. a contract at time t promising that we *will* buy the bond at a pre-specified price p^f after days. This will not require any cash disbursement until the settlement time *t* + arrives, while yielding a gain or loss given the way the market prices move until that time. Here, the forward price p^f and the spot price *P* will not be identical. The underlying asset being the same, we can still anticipate quite *similar* profits and losses from the two positions.

This illustrates one of the basic premises of financial engineering. Namely that as much as possible, one should operate by taking positions that do not require new funding.

Arbitrage

The notion of arbitrage is central to financial engineering. It means two different things, depending on whether we look at it from the point of view of market *practice* or from the *theory* of financial engineering.

We begin with the definition used in the theory of financial engineering. In that context, we say that given a set of financial instruments and their prices, *{PI, P2. . ., Pk},* there is no arbitrage opportunity if a portfolio that costs nothing to assemble now, with a non-negative return in the future is ruled out. A portfolio with negative price and zero future return should not exist either.

If prices *Pi* have this characteristic, we say that they are *arbitrage-free.* In a sense, arbitrage free prices represent the *fair* market value of the underlying instruments. One should not realize gains without taking some risk and without some initial investment. Many arguments in later chapters will be based on the no-arbitrage principle.

In market practice, the term "arbitrage," or "arb," has a different meaning. In fact, "arb" represents a position that *has* risks, a position that *may* lose money but is still highly likely to yield a "high" profit.

The Syndication Process

A discussion of the syndication process will be useful. Several contract design and pricing issues faced by a financial engineer may relate to the dynamics of the syndication process. Stocks, bonds, and other instruments are not sold to investors in the primary market the way, say, cars or food are sold. The selling process may take a few weeks and has its own wrinkles; these may end up being quite important for a financial engineer.

Selling Securities in the' Primary Market

The following gives an indicative time table for a syndication process. Such time tables show variations from one instrument to another. Even in the same sector, the timing may be very different from one issuer to another, depending on the market psychology at that time. The process described gives an example. The example deals with a Eurobond issue. For syndicated loans, for facilities, and especially for IPOs, the process may be significantly different, although the basic ideas will be similar.

1. The week of D-14: Manager is chosen, mandate is given. Issue strategy is determined.

   Documentation begins. 2. The week ofD-7: Documentation completed. Co-managers are determined.

3. D-Day: "Launch" date. Sending faxes to underwriter and selling group members. Issue is published in the press.

4. D+8: Preliminary allotment done by lead manager.
5. D+9: Pricing day.
6 D+ 10: Offering day. Allotment faxes are sent to group members.
7. D+24: Payment day. Syndicate members make payments.

In other markets, important deviations in terms of both timing, and procedure may occur during actual syndication processes. But overall, the important steps of the process are as shown in this simple example.

Syndication of a Bond versus a Syndicated Loan

We can compare a bond issue with processing a syndicated loan. There are some differences. Syndicated loans are instruments that are in banking books or credit departments of banks. The follow-up and risk management is done by the banking credit departments with methodologies similar to standard loans. For example, information in the offering circular is not very important.

Bonds, on the other hand, are handled by investment or in trading books, and the analysis and information in the circular are taken seriously. Documentation differences are major.

The syndicated loan tries to maintain a relationship between the bank and its client through the agent. But in the bond issue, the relationship between the lender and the borrower is much more distant. Hence, this type of borrowing is available only to good names with good credit standing. (Banks have to continuously follow lesser names to stay aware of any deterioration of credit conditions.) The maturities can also be very different.

You solve the following problems at home for your better understanding.

Exercises

1. Suppose the quoted swap rate is 5.06/5.10. Calculate the amount of fixed payments for a fixed payer swap for the currencies below in a 100 million swap.
   - USD.
   - EUR.

   Now calculate the amount of fixed payments *for* a fixed receiver swap *for* the currencies below in a 100 million swap.
   - JPY
   - GBP.

.2. Suppose the following stock prices for GE and Honeywell were observed before any talk *of* merger between the two institutions:

Honeywell (HON) 27.80

General electric (GE) 53.98

Also suppose you "know" somehow that GE will offer 1,055 GE shares for, each, Honeywell share during any merger talks.

a. What type of "arbitrage" position would you take to benefit from this news?

b. Do you need to deposit any of your funds to take this position?

c. Do you need to and can you borrow funds for this position?

d. Is this a true arbitrage in the academic sense of the word?

e. What (if any) risks arc you taking?

3. Read the market example below and answer the following questions that relate to it.

"Proprietary dealers are betting that Euribor, the proposed continental European-based euro money market rate, will fix above the Euro BBA Libor alternative. The arbitrage itself is relatively straightforward. Tire proprietary dealer buys the Life September 1999 Euromark contract and sells tire Matif September 1999 Pibor contract at roughly net zero cost. As the Life contract will be referenced to Euro BBA Libor and tire Matif contract will be indexed to Euribor, the trader in effect receives Euribor and pays Euro BBA Libor.

The strategy is based 011 the view that Euribor will generally set higher than Euro BBA Libor. Proprietary dealers last week argued that Euribor would be based on quotes from 57 different banks, some of which, they claimed, would' have lower credit ratings than the eight Libor banks. In contrast, Euro BBA Libor will be calculated from quotes from just 16 institutions. (From IFR, December 18, 1998)

a. Show the positions of the proprietary dealers using position diagrams.

b. In particular what on the horizontal axis of these diagrams? What on the vertical axis?

c. How would the profits of the "prop" dealers be affected at expiration. If in the meantime there was a dramatic lowering of all European interest rates due, say, to a sudden recession?

INTERNATIONAL MONETARY SYSTEM (IMS)

Objectives

- To distinguish between free float, managed float and fixed exchange rate system
- To identify the categories of central bank intervention
- To identify the post war international monetary system and its role.

Dear friends, multinational corporations operate in a global market, buying/selling/producing in many different countries. For example, GM sells cars in 200 countries, produces cars in 50 countries, so it has to deal with close to 200 currencies. What are the mechanics of how currency and capital flows internationally?

International Monetary System - Institutional framework within which

1. International payments are made
2. Movements of capital are accommodated
3. Ex-rates are determined

An international monetary system is required to facilitate international trade, business, travel, investment, foreign aid, etc. For domestic economy, we would study Money and Banking to understand the domestic institutional framework of money, monetary policy, central banking, commercial banking, check-clearing, etc. To understand the flow of international capital/currency we study the IMS. IMS - complex system of international arrangements, rules, institutions, policies in regard to ex-rates, international payments, capital flows. IMS has evolved over time as international trade/finance/business have changed, as technology has improved, as political dynamics change, etc. Example: evolution of the European Union and the Euro currency impacts the IMS. Before we go to the technalities of IMS, let's understand little bit the origin of this body.

History of The Ims

Bimetallism(pre-1875)

Commodity money system using both silver and gold (precious metals) for international payments (and for domestic currency). Why silver and gold? (Intrinsic Value, Portable, Recognizable, Homogenous/Divisible, Durable/Non-perishable). Why two metals and not one (silver standard or gold standard vs. bimetallism)? Some countries' currencies in certain periods were on either the gold standard (British pound) or the silver standard (German DM) and some on a bimetallic (French franc). Pound/Franc ex-rate was determined by the gold content of the two currencies. Franc/DM was determined by the silver content of the two currencies. Pound (gold) / DM (silver) rate was determined by their ex-rates against the Franc.

Under a bimetallic standard (or any time when more than one type of currency is acceptable for payment), countries would experience "Gresham's Law" which is when "bad" money drives out "good" money.

The more desirable, superior form of money is hoarded and withdrawn from circulation, and people use the inferior or bad money to make payments. The bad money circulates, the good money is hoarded. Under a bimetallic standard the silver/gold ratio was fixed at a legal rate. When the market rate for silver/gold differed substantially from the legal rate, one metal would be overvalued and one would be undervalued. People would circulate the undervalued (bad) money and hoard the overvalued (good) money.

You can just see this example:

a) From 1837-1860 the legal silver/gold ratio was 16/1 and the market ratio was 15.5/1. One oz of gold would trade for 15.5 oz. of silver in the market, but one oz of gold would trade for 16 oz of silver at the legal/official rate. Gold was overvalued at the legal rate, silver was undervalued. Gold circulated and silver was hoarded (or not minted into coins), putting the US on what was effectively a gold standard.

b) France went from a bimetallic standard to effectively a gold standard after the discovery of gold in US and Australia in the 1800s. The fixed legal ratio was out of line with the true market rate. Gold became more abundant, lowering its scarcity/value, silver became more valuable. Only gold circulated as a medium of exchange.

Classical Gold Standard (1875-wwi)

after that for about 40 years most of the world was on an international gold standard, ended with WWI when most countries went off gold standard. London was the financial center of the world, most advanced economy with the most international trade. Then you see what has happened.

Gold Standard exists when most countries:

1. Use gold coins as the primary medium of exchange
2. Have a fixed ex-rate between ounce of gold and currency
3. Allow unrestricted gold flows - gold can be exported or imported freely.
4. Banknotes had to be backed with gold to assure full convertibility to gold.
5. Domestic money stock had to rise and fall with gold flows.

   Under a gold standard, ex-rates would be kept in line by cross-country gold flows. Any mis-alignment of ex-rates would be corrected by gold flows. Payments could in effect be made by either gold or banknotes. If market ex- rates ever deviated from the official ex-rate, it would be cheaper to pay in gold than in banknotes. We can see one more example for your understanding.

Example: suppose that the UK Pound is pegged to gold at: 6 Pound/oz., and the franc is pegged at 12 FF/oz, then the official ex-rate should be 2FF/Pound. If the market rate is 1.8FF/Pound, then the pound is undervalued in the market (one pound should buy 2 FF, it only buys 1.8 FF). Arbitrage would re-align the ex-rate.

1. Take £500 and buy 83.33 oz of gold (£500/6). 2. Buy 1000 FF (83.33 oz x 12) 3. Sell 1000 FF for £533.33 (FF1000/1.8FF/£)

Gold would move from UK to France, which would depreciate the FF and appreciate the £, ex-rate would be restored at 2FF/£.

And at the same time under gold standard, international balance of payments gets corrected automatically. Suppose that UK has a trade surplus (X > M) with France (M > X) which has a trade deficit. UK sold more to France than it bought, France bought more from UK than it sold, which brings about a flow of gold from France to UK. The increased (decreased) gold in UK (France) brings about inflation in UK and deflation in France. As time goes on, Exports from UK will fall because British prices are now higher, Imports will rise because French prices are lower. The trade surplus of UK will fall and France's deficit will fall. Market forces automatically correct trade deficits/surpluses; this adjustment mechanism is known as the price-specie-flow mechanism.

Advantages of Gold Standard

1. Ultimate hedge against inflation. Because of its fixed supply, gold standard creates price level stability, eliminates abuse by central bank/hyperinflation.
2. Automatic adjustment in Balance of payments due to price-specie-flow mechanism.

Disadvantages of Gold Standard

1. Possible deflationary pressure. With a fixed supply of gold (fixed money supply), output growth would lead to deflation.
2. An international gold standard has no commitment mechanism, or enforcement mechanism, to keep countries on the gold standard if they decide to abandon gold.

Interwar Period: 1915-1944

When WWI started, countries abandoned the gold standard, suspended redemption of banknotes for gold, and imposed embargoes on gold exports (no gold could leave the country). After the war, hyperinflationary finance followed in many countries such as Germany, Austria, Hungary, Poland, etc. Price level increased in Germany by 1 trillion times. **Why hyperinflation then?? What are the costs of inflation??**

US (1919), UK(1925), Switzerland, France returned to the gold standard during the 1920s. However, most central banks engaged in a process called "sterilization" where they would counteract and neutralize the price-specie-flow adjustment mechanism. Central banks would match inflows of gold with reductions in the domestic MS, and outflows of gold with increases in MS, so that the domestic price level wouldn't change. Adjustment mechanism would not be allowed to work. If the US had a trade surplus, there would be a gold inflow which should have increased US prices, making US less competitive. Sterilization would involve contractionary monetary policy to offset the gold inflow.

In the 1930s, what was left of the gold standard faded - countries started abandoning the gold standard, mostly because of the Great Depression, bank failures, stock market crashes. Started in US, spread to the rest of the world. Also, escalating protectionism (trade wars) brought int'l trade to a standstill. (Smoot-Hawley Act in 1930), slowing int'l gold flows. US went off gold in 1933, France lasted until 1936.

Between WWI and WWII, the gold standard never really worked, it never received the full commitment of countries. Also, it was period of political instability, the Great Depressions, etc. So there really was no stable, coherent IMS, with adverse effects on int'l trade, finance and investment.

Bretton Woods System: 1945-1972

At the end of WWII, 44 countries nations met at Bretton Woods, N.H. to develop a postwar IMS. The IMF and the World Bank were created as part of a comprehensive plan to start a new IMS. The IMF was to supervise the rules and policies of a new fixed ex-rate regime; the World Bank was responsible for financing development projects for developing countries (power plants, roads, infrastructure investments).

IMS established by Bretton Woods was a **dollar-based, gold-exchange standard of fixed exchange rates.** The US dollar was pegged to gold at a fixed price of $35/ounce, and then each currency had a fixed ex-rate with the $.

Examples: $2.80/£, 4.2DM/$, 3.5FF/$, etc.

Each country was supposed to maintain the fixed rate within 1% of the agreed upon rate, by buying/selling currency. To increase the foreign exchange value of DM, the central bank would buy back DMs with $, to decrease the value of DM it would sell DMs for $. US $ was convertible to gold, the other currencies were not. Countries held $ and gold for IMS payments. A country with a "fundamental disequilibrium" could be allowed to change its fixed rate with the $.

Advantages of Gold-Exchange System/ Bretton Woods in SR:

1. Economizes on scarce resources (gold) by allowing foreign reserves ($s) to be used for IMS payments. Easier to transfer dollars vs shipping gold overseas under pure gold standard.
2. By holding $ instead of gold as reserves, foreign central banks can earn interest vs. non-interest bearing gold.
3. Ex-rate stability reduced currency risk, provided a stable IMS, and facilitated international trade and investment, led to strong economic growth around the world in 50s and 60s.

In long run, Bretton Woods (gold-exchange system) was unstable. There was no way to:

1. Devalue the reserve currency ($) even when it was overvalued or
2. Force a country to revise its ex-rate upward. A country could agree, or be pressured into devaluation, but there was no way to "revalue" a currency upward (appreciate through contractionary policy). In the 1960s, US pursued expansionary monetary policy (printed money) to reduce unemployment, resulting in the dollar being overvalued and foreign currencies being undervalued according to the fixed ex-rate system. There was no way to devalue the $, and other countries were not willing to revalue their ex-rates upward. Why?

Bretton Woods started to collapse in 1971, temporary measure (Smithsonian Agreement) didn't work, fixed ex-rate regime was abandoned in 1973. Also, Nixon put wage and price controls went into effect in 1971, were then lifted, first oil shock started in 73 (Arab oil embargo after Nixon gave $2.5B to Israel after Egypt attacked), oil prices doubled, no way to stabilize the dollar. 1973- Fixed ex-rates/Bretton Woods were abandoned.

Flexible Exchange Rates: 1973-present

IMF members met in Jamaica in 1976 to agree to a new IMS including:

a. Flexible ex-rates allowed, central banks could intervene in currency markets. (Under fixed ex-rates, you lose control over your monetary policy. Monetary policy must be committed to maintaining the fixed ex-rate, and cannot be used to pursue other macroeconomic goals)
b. Gold was abandoned as a reserve asset.
c. Developing countries were to get more assistance from IMF.

IMF was to provide assistance to countries facing BP/currency difficulties. (Brazil is an example). IMF provides grants and loans to countries with problems under the conditions that they follow IMF's policy prescriptions - "strings attached to aid." Reduced budget deficits, reduced govt. spending/cutting subsidies, contractionary monetary policy, i.e. responsible fiscal and monetary policies.

Advantages of Flexible Ex-Rates

a. Countries have control over monetary policy
b. A true market value is established for currency, fluctuates daily to reflect market forces of S and D.
c. Flexible ex-rates maintain BP equilibrium. Example: U.S. has trade deficit, M>X, excess dollars in world currency markets, $ depreciates, £ appreciates, US exports will go up, restore trade balance.

Disadvantages

a. More Volatility, see page 34. MNCs must be concerned about currency risk.
b. Potential abuse by central bank, reckless monetary expansion.

But here also major currencies like $, £ Yen, etc. are freely floating ex-rates, changing daily to reflect market forces. Most of the rest of the world is under some type of system of "pegged ex-rates" or "managed floating", where central bank intervention is required to maintain a certain level of ex-rates. One system results in trading 1:1 with the dollar (Panama, Bahamas, Belize 2:1, and Liberia), other systems trade within a certain band (range). Currencies pegged to $, FF, SDRs, others. 41 are independently floating, no pegging or targeting. 42 have "managed floating" systems that combine market forces with pegging.

European Monetary System has been replaced by the Euro, the single currency in Europe. (1 ECU = 1 Euro) To qualify, countries had to meet certain economic criteria

1. Deficits/GDP less than 3%,
2. Price level stability - low and stable inflation, etc. Of the 15 countries in the European Union, three countries decided not to join (UK, Denmark, and Sweden).

You see some of the recent happening:
As of Jan 1 1999:

1. The 12 countries fixed their ex-rates against each other and against the Euro and
2. The Euro became a unit of account. For example, 3.35FF/DM. 6.55 FF/Euro. FF and DM will float against the $, £ and Yen, but will be fixed against each other and against the Euro. Fixed ex- rate system for the 11 countries.

Euro currency (euro as a medium of exchange) started to circulate on Jan. 1, 2002. Old currency and Euros BOTH circulated for the first 6 months, then old currency was taken out of circulation and only Euros now exist.

Changes

1. Stores now quote prices in Euros.
2. Payment in Euros can be made with charge cards and checking accounts
3. Euro currencies options are now traded
4. Stock prices/indexes are quoted in Euros.
5. European Central Bank (ECB) established to conduct monetary policy in Europe. Governing Council made up of 12 ECB governors, one from each country, and 6 members Executive Board.

See the $/€ ex-rate on p. 42. Started at $1.18/€ in 1999, fell to $0.83/€ in 2000 (why?), then started to increase; now it's about $1.08 (9/2003).

Main Advantages of Euro (€)

1. Significant reduction in transaction costs for consumers, businesses, governments, etc. (estimated to be .4% of European GDP, about $50B!)"European Saying: If you travel through all 15 countries and exchange money in each country but don't spend it, you end up with 1/2 of the original amount!
2. Elimination of currency risk, which will save companies hedging costs
3. Promote corporate restructuring via M&A activity (mergers and acquisitions), encourage optimal business location decisions.

Main Disadvantage of Euro

Loss of control over domestic monetary policy and exchange rate determination.

Suppose that the Finnish economy is not well-diversified, and is dependent on exports of paper/pulp products, it might be prone to "asymmetric shocks" to its economy. If there is a sudden drop in world paper/pulp prices, the Finnish economy could go into recession, unemployment could increase. If independent, Finland could use monetary stimulus to lower interest rates and lower the value of its currency, to stimulate the domestic economy and increase exports. As part of EU, Finland no longer has those options; it is under the EU Central Bank, which will probably not adjust policy for the Euro zone to accommodate Finland's recession. Finland may have a prolonged recession. There are also limits to the degree of fiscal stimulus through tax cuts, since budget deficits cannot exceed

3% of GDP, a requirement to maintain membership in EMU (to discourage irresponsible fiscal behavior).

General Consensus: Euro has been a success, and will likely emerge as the second global currency, with the Yen as a junior partner. The success of the Euro may encourage other areas to explore cooperative monetary arrangements (Asia, S. America) and three world currencies at some point (¥, €, $).

For your additional information you can see official Euro web site (in English) at: http://europa.eu.int/euro/entry.html or the website of the European Central Bank at: http://www.ecb.int

We are meeting in the next class. If you have any problem, I will try to explain you again.

Bye.

FOREIGN EXCHANGE MARKETS

Objective

- To describe the organization of foreign exchange market and distinguish between spot and forward rate.
- To distinguish between different methods of foreign exchange quotation and convert from one method of quotation to another method.
- To read and understand foreign currency quotation as they appear in the Wall Street journals.

Dear friends, let's come today to the foreign exchange market. What do you feel about this topic- "FOREIGN EXCHANGE (FX) MARKET"? Coming to the point-

money represents purchasing power (JS Mill: "There is nothing more insignificant than money"), but usually only in one country. ¥, £, or € have no purchasing power in U.S. Exchanging one currency for another takes place in the FX market (converting purchasing power from one currency into another).

FX is world's largest financial market in world. $1.2T/day vs. $5T/year for all publicly traded stock in U.S. ($20B/day in US stocks vs. $1200B/day in FX). FX market trading has decreased in recent years.

FX traded 24/7/365 hours in a year you should understand that it does not happened only in one particular place or location. Three major areas: Australasia (Sydney, Tokyo, Hong Kong, Singapore), Europe (London, Paris, Amsterdam, Frankfurt) and N. America (NY, Montreal, Toronto, SF, Chicago, LA, etc.) are the larger market for FX trading. Due to time zone differences, trading takes place 24 hours/day.

Most trading rooms operate 9-12 hour days. Trading volume is high when the N. America and Europe markets overlap (early in the day in U.S.) and late in the day in Asia when the European market is opening.

Two largest trading centers: 33% of daily trading volume in London/UK ($504B daily), and 17% in U.S. ($254B daily).

FX actually covers spot currency markets, forward currency markets, currency options, currency swaps, currency futures, foreign trade financing and credit arrangement, etc. Here we will discuss about spot and forward markets only.

Function and Structure of Fx Market

FX markets are part of Commercial Banking activities, assisting corporate clients/MNCs to conduct international commerce. Banks provide the service of buying/selling foreign currency for commercial customers, e.g. importers who are buying foreign products and need to buy foreign currency with $, or exporters who are receiving foreign currency and need to sell foreign currency for $

FX is an OTC (over-the-counter) market, like NASDAQ. But you can ask me of how OTC differs from non-OTC.

FX OTC market is an international network of bank currency traders, non-bank dealers, FX brokers, linked by computers, phone lines, telex machines, automated quotation systems, etc. The communication system of FX dealers is extremely advanced, sophisticated and reliable in comparison to other exchange markets.

Fx Market Participants

Two levels: Wholesale (Inter-bank, 87% of trading volume) and Retail (Client market, 13% of market). Why so much wholesale? Because such wholesale trading is speculative trading (trying to correctly judge the direction of currency values) or arbitrage trading what I mean exploiting ex-rate discrepancies between dealers. Currency trading is a profit center for large banks.

1 **Intl Banks and Bank Customers.** 100-200 large commercial banks worldwide provide the core of the FX market and actively participate, and "make a market" in FX, trading FX on behalf of bank customers (MNCs, money managers, exporters, importers, private traders).

2. **Nonbank Dealers.** Wholesale currency traders who are NOT commercial banks, e.g. Investment banks (Solomon Smith Barney, M-L, JP Morgan, Goldman Sachs, etc.), who establish their own trading centers to trade directly in the FX market, and account for 28% of the inter-bank (wholesale) volume.

3. **Fx Brokers.** Brokers/intermediaries who track quotes offered by many dealers in the global market, and then match buyers and sellers for a fee (bid/ask spread), and "make a market," without taking a position themselves (no currency inventory). More and more trading (50-70%) now takes place through automated electronic trading systems, making the role of FX brokers unnecessary. This trend will continue as electronic trading becomes more advanced.

4. **Central Banks.** If a country has a fixed ex-rate (Argentina until recently, Hong Kong, Belize), or a pegged rate (China), the central bank (or currency board) has to make regular interventions to support the fixed/pegged ex-rate.

For an Example: China's pegged rate is Y8.28/$. If the ex-rate goes toward 8.4 (8.2), Yuan is depreciating (appreciating), central bank must buy (sell) Yuan to strengthen/appreciate (weaken/depreciate) the Yuan. Buying (selling) pressure will increase (decrease) the value of the currency.

Even under floating rates, central banks may intervene in FX market, to influence the domestic (or foreign trading partner's) currency for policy goals.

Another example which can be quoted here is: In mid-80s, dollar was strong, US manufacturers complained to Reagan admin. Why? Solution?
Japan tries to keep the US $ strong. Why?

You will come to know this answer from the following explanation.

Correspondent Banking Relationships

Wholesale inter-bank FX trading among commercial banks, who maintain demand deposits (corresponding banking accounts) with each other, creating a wholesale banking network. ""U.S. importer buying merchandise from a Dutch exporter invoiced in Euros at an agreed upon price of €512,100. Correspondent banking systems handles the currency trading. Importer contacts U.S. bank to get an ex-rate quote to buy €512,100 @ €1.0242/$ (Note that: ex-rates typically quoted to 4 decimal places), or $500,000. Funds get transferred between U.S. Bank and its correspondent bank in Europe (EZ Bank), where the Dutch Exporter has an account. Thus, the US Importer and Dutch Exporter handle the currency transfer through their respective banks, who have a correspondent banking relationship.

The transfer of funds internationally is facilitated by international clearinghouse services like SWIFT (Society for Worldwide Inter-bank Financial Telecommunications), CHIPS (Clearing House Inter-bank Payments System, part of the FRS) and ECHO (Exchange Clearing House Limited). SWIFT and CHIPS also provide check-clearing services, account transfers, wire transfers, etc. ECHO is exclusively for FX.

Fx Spot Market

Cash (or spot) market for currency, involves immediate delivery (within one or two days), represents 33% of total FX market. Spot rates (S) can be quoted two ways. Ex-rate is just a ratio of two currencies, can be expressed two ways: S = $/£, or S = £/$. (1/x key on calculator).

When the dollar is on the top of the ratio, S = $ / £, or S($/£) = $1.5272/£ this is called:

a. Direct quote (priced in dollars)

b. U.S. $ Equivalent

c. American terms

When the dollar is on the bottom of the ratio S = £ / $ or S(£/$) = £0.6548/$, this is called:

a. Indirect Quote

b. European terms

c. Currency PER US $

Spot rates are reported both ways. You can refer to the wall street journal for proper understanding.

Most currencies are priced and traded against the $ (90% world currency market involves the dollar on one side of transaction). General rule: All currencies are generally quoted in European terms, Indirect quote, e.g., ¥118/$, Mex Pesos 9.8130 / $, EXCEPT British pound (£) and former British colonies (Australia, NZ and Ireland), which are quoted in American terms, Direct, e.g. $1.5272/ £. Reason: pre-1971, British pound and currencies based on the £ were NON-decimal, so it was more convenient to report Spot Rate as $/ £, a practice that continues until today. When Euro was introduced, it was decided it would also be quoted as a direct quote, $1.1139/€. HINT for you is: Always pay close attention to how currency is quoted, direct or indirect."S($/£) or S(£/$). Reciprocal of one another.""S = $1.5272 / £""S = 1/1.5578 = £0.6548 / $."

Calculator: Use 1/x key (HP: 1/x is on the Divide key)

Bid-ask Spread

Bid/Ask Spread provides a commission/brokerage fee for currency traders/brokers. The FX currency trader/bank will BUY FX for inventory at the BID price and SELL FX at the ASK price. BID price is always LOWER than the ASK price. Dealers buys low, sells high.

Example: S(BID) = £0.6548 / $
S(ASK) = £0.6550 / $

Dealer will pay £0.6548 for a $, and sell the $ for £0.6550, spread is the profit margin (£0.6421 - £0.6419 = £0.0002). Most wholesale, standard-size transactions are for $10m or more, so the spread generates profits even though it is very low, ($10m x £0.0002/$ = £2000 profit, or about $3000). Retail bid-ask spreads are higher, more profitable than wholesale spreads, to cover the fixed cost of a transaction that applies to even small currency trades.

Currency trading rooms are set up so that individual traders specialize in one currency and trade it against the $: ¥, €, £, SF, etc. There is also a "cross-rate" desk for trades NOT involving the $, e.g., ¥ / €. Traders might make 400 trades per day. See story "Young Traders Run Currency Markets" on pages 84-85. " ""CROSS EX-RATES""Cross Ex-Rate is an ex-rate that does NOT involve $, e.g. €/£. "See this example: S(€/£) = S($/£) / S($/€) = S($/£) x S(€/$), so S = €/£

(The $s cancel, leaving €/£).""From Exhibit 4.4: S ($/£) = 1.5272 and S ($/€) = .9764

So $\frac{S(\$/£)}{S(\$/€)}$ = 1.5272 / .9764 = €1.5641/£.

For 9 currencies there would be (9 x 8)/2 = 36 cross currency rates.

"CROSS-RATE TRADING""To simplify trading, most trades goes through the $, even for a non-dollar "currency against currency" trade, e.g. trading £s for SF. Trade: £s for $s, $s for SFs instead of £s directly for SFs. Reason: Assume there are 9 major currencies including $. There are 8 trading desks, each quoting a rate using US $. If each 9 currencies were also traded against each other, there would have to be 36 trading desks [((N-1) x N) / 2] instead of 8 to have all combinations of two currencies, or traders would have to deal in more than one currency, which would be too complicated and complex.

Cross trades (currency against currency) are handled at a special cross-rate desk, going through the dollar ex-rate for each currency."

Triangular Arbitrage"

This means exploiting FX discrepancies using 3 currencies, to make risk less profits. "Example: $1 to £, £ to €, and € back to $. Supposed you start with a $1, end up with $1.01.""$/£ x £/€ x €/$ should = 1

1.529 x .639 x 1.024 H•1.000

When it does not equal 1, there is Triangular Arbitrage."*Example:* 1.529 x .639 x 1.0337 = 1.01"$1 turns into $1.01, $1m turns into $1.01m = $10,000 profit.

Reason: $ is overvalued to the € ($1 should buy only €1.024, but actually buys €1.0337, about 1% overvalued).”” Arbitrage profits from triangular arbitrage would typically be: a) small, b) infrequent and c) temporary. “Picking up dimes with a bull-dozer.” Efficient market hypothesis. Law of One Price. Price Equalization Principle.

Forward Market

Contract settled today for future delivery/receipt of FX. Agree today on P (ex-rate) and Q, future settlement in 1, 3, 6, 9, 12 months, 2, 5, 10 years, etc. Forward rates are available for most major currencies at most maturities.

Compared to the spot rate, FX is usually trading at either a Forward Discount (currency is expected to depreciate) or Forward Premium (currency is expected to appreciate). Which FX is selling at discount/premium?

Notation in Book: S ($/SF) = Spot rate
F_1 ($/SF), F_3 ($/SF), F_6($/SF) are 1, 3, and 6 month forward rates.

S($/SF) = .6653

F_1($/SF) = .6660

F_3($/SF) = .6670

F_6($/SF) = .6684 (% CHG = .466% x 2 = .932% annual)

SF is selling at a forward premium of about 1% (annual), dollar is selling at forward discount of about 1% per year. SF is expected to *appreciate*, dollar is expected to *depreciate*, and those expectations are already being priced in the forward market.
Calculation of Forward Premium/Discount in %
S ($/SF) = .6653
F_6($/SF) = .6684

Forward Premium/Discount

[(F - S) / S] x 100%CHG = [(.6684 - .6653) / .6653] x 100 = .466% x 2 = .932%

HP 10B: .6653 / INPUT / .6684 / Yellow Key / % (.466 in Display x 2 = .932%)

Two groups of participants for Forward/Futures Markets:

1. Hedgers - Investors/companies who are exposed to currency risk. Allows MNCs to control and manage revenues and payments
2. Speculators - Pure speculative position on currency rates, gambling on ex-rates. Go long if you expect future S > F_6, go short if you expect future S < F_6.

Example of Speculative Forward Position

F_3($/SF) = $.6670 but S_3 will probably not be exactly $.6670, but you can lock in now to buy or sell at $.6670/SF. If you think S_3 will be > $.6670, you go LONG, buy SF forward, and lock in at .6670, expecting S >$.6670 in 3 months. Hopefully, you buy at F_3 = $.6670 (guaranteed) and sell at S > $.6670 in 3 months, make profit of (S_3 - $.6670).

If you think S_3 < $.6670, you go SHORT, sell SF forward, and lock in to sell at $.6670, if S_3 < $.6670, you can buy at < $.6670 and sell at $.6670, and make money.

Example. SF trader thinks that the S_3 < F_3 $.6670, so takes a SHORT position, sells SF5,000,000 forward in 3 months against the $, based on belief that the SF will ***appreciate less*** than expected, dollar will ***depreciate less*** than expected. If S_3 < F_3 $.6670, trader will make money. If S_3 = $.6600, trader can buy SF at $.6600 and sell SF at $.6670, make money ($.6670 - $.6600 = $.0070/SF x SF 5, 000,000 = $35,000 profit).
If trader is wrong and S_3 > F_3, then they lose money. If S_3 = $.6700 then they have to buy at $.6700 and sell at $.6670 for a loss of -$.0030 per SF, x SF 5, 000,000 = -$15,000.

Point: a) Hedgers are using the forward ex-rate markets to manage, control or eliminate currency risk, because they have some personal or business interest in the outcome of currency movements. Forward markets are like insurance markets for the hedgers, e.g., importers, exporters, foreign investors, etc.

b. Speculators are using the forward ex-rate markets to take pure speculative positions on currency changes, pure gambling, with no business interest in currency changes.

I hope you will be able to grasp this concept so that we can move to the next step comfortably. Okey. Thank you all.

INTRODUCTIONS TO DERIVATIVES

Objectives

After completion of this lesson you will be able to understand what is derivative and how the financial instruments are used in the capital markets of the world.

So, all of you are geared up for the next lesson. You must have come across the term "Derivatives" quite often but you must be wondering what it means. "Derivatives" is a very interesting topic and I have made the lesson in such a manner that it becomes easier for you to understand.

Derivative is an instrument, whose value is derived from the value and characteristics of the underlying asset. It is basically a contract between two parties that specifies conditions in particular, dates and the resulting values of underlying variables under which payments or **pay-offs** are to be made between the parties.

Just see as an example, social security is a derivative which requires a series of payments from an individual to the government before age 65, and payoffs after age 65 from the government to the individual as long as the individual remains alive. In this case, the payoffs occur at predefined dates and depend on the individual's survival. Anyone who has ever taken out a mortgage with a prepayment privilege has perhaps unwittingly dabbled in derivatives. Let us take a more dramatic example, earthquake insurance is a derivative in which an individual makes regular annual payments in exchange for a potentially much larger payoff from the insurance company should an earthquake destroy his property. Derivatives are also known as **contingent claims** since their payoffs are "contingent" upon the outcome of an underlying variable.

Of course derivatives have long existed with specific events or commodity prices as underlying variables. The big explosion of interest in derivatives, however, occurred only after purely financial derivatives with stock prices, stock indexes, foreign exchange rates, bond prices and interest rates became the variables determining the size of payoffs. Historians searching for a starting date might look to 1972, the formation of the International Monetary Market (IMM), a division of the Chicago Mercantile Exchange (CME), or April 1973, the opening of the Chicago Board Options Exchange (CBOE), the first modern exchanges to trade financial derivatives.

Speaking philosophically and very much in the spirit of this **paper**, interpreting something as a derivative depends on one's point of view. For example, common stock is usually considered an asset that might underlie a derivative, but not a derivative itself. Yet, if the payoff from stock is considered dependent on some other underlying variable such as the operating income of the associated firm, then the stock itself is being interpreted as a derivative. Whether or not it pays to make this interpretation depends on the particular purpose at hand. To take a classic example from another field, for some purposes it is best to think of the sun as fixed in space and the earth as rotating around it; for others it pays to adopt the Aristotelian perspective of the earth fixed in space with the sun rotating about it.

If You Recall Correctly there are Four Main Financial Markets. They are

Derivative Markets: forward contracts, futures, options, future options, swaps, etc. Generally, a derivative security "derives" its value from the price movements in some underlying commodity, currency, common stock, stock index, T-bill, interest rate, etc. It is like a "side bet."

Why do derivative markets exist? Largely to facilitate hedging, the derivative markets are largely **insurance markets.** We saw in the last few chapters how interest rate risk played an important role in the S&L crisis. We also studied the potential adverse effects of currency risk on an international firm's profits. Firms, like individuals, are "risk averse" and would like to protect themselves against the three main types of risk that businesses face: **PRICE RISK, CURRENCY RISK and INTEREST RATE RISK.**

In this chapter, we look at futures contracts, and study the important role that they play in risk management and risk-sharing by allowing firms to hedge risk. The text focuses specifically on financial derivatives (used to hedge interest rate risk), but we will consider a broader coverage of futures.

Spot Market vs. Forward/Futures Market: In the spot (cash) market, buyers and sellers agree on Price (P) and Quantity (Q) for immediate delivery (or within a few days). *Examples:* Ford buys 1m German marks in the spot market for currency, or it buys 1m pounds of steel in the cash market for steel. Or Mars Candy Company buys 1m pounds of sugar in the cash market. Northwest Airlines buys 500,000 gallons of gasoline in the spot market.

Forward Contracts: Private contracts between two parties (buyer and seller) agreeing to an exchange in the future. Buyer and seller agree on Price and Quantity today, for delivery sometime in the future (one month, one year, ten years). Forward contracts are private contracts, and are therefore not marketable securities, there is no secondary market, e.g. like the difference between a bank loan (not marketable) and a bond (marketable). We studied forward rates and forward contracts for foreign exchange in the previous lesson.

Example: Jolly Green Giant Co., or Pepsi Cola, enters into a forward contract in May to purchase corn at harvest time in October, at a guaranteed price, from various farmers for their entire crop. Advantage: buyer (company) and the seller (farmer) have a guaranteed price. They are now protected from price swings in corn, they have eliminated price risk completely by hedging their position, locking in a price with a forward contract.

Example: GM enters into a forward contract for British pounds with Bank One, to either buy pounds or sell pounds, in six months at a guaranteed ex-rate. By locking in, GM has hedged currency risk.

Advantage of Forward Contracts: they are very flexible can be customized to the needs of the parties.

Disadvantages of Forward Contracts

1. There is not a liquid market for forward contracts, no secondary market. Might be hard to match up the two parties to the transaction.
2. High default risk. No outside party guaranteeing the transaction, like there is in the futures market.
3. Requires actual delivery to complete the contract.

Futures Contracts are the same in principle as a forward contract, where two parties (buyer and seller) agree to trade/exchange something (corn, oil, gold, Tbills, Yen) in the future (one week, one month, one year, ten years), but they agree on P and Q now, for future delivery, using a futures contract from a futures exchange - an organized market for trading futures contracts.

Advantages of futures contracts over forward contracts

1. Liquid market, lots of buyers and sellers at organized exchanges all over the world (see handout).
2. Active secondary market. Contracts may trade hands many times before expiration.
3. Minimal risk - the futures exchange requires an initial margin requirement to open a position and they enforce daily settlement of all gains and losses to avoid default. There is a maximum price movement, called the daily limit, to minimize large losses. Example: daily price limit for wheat futures contracts is 20 cents per bushel, trading stops for the day.
4. Cash settlement for most futures contracts, instead of settlement in the actual commodity.
5. You can close out your account any time by taking an offsetting position. If your original position is to buy (go long) a futures contract, you can subsequently sell (go short) to close out your position, and vice versa. You are basically agreeing to sell the contract to yourself, so you can cash out without having to make or receive delivery.

Disadvantages of Futures Contracts Over Forward Contract

1. Less flexible, since futures contracts are for fixed, standard amounts, e.g. corn futures contracts are for 5,000 bushels per contract.
2. Expiration dates are fixed, e.g. Jan, March, May, July, September, and December for corn contracts, so there are only six delivery days per year.

Example: Suppose a corn farmer expects a yield of 7,500 bushels of corn at harvest next October, and wants to use a futures contract to hedge commodity price risk by taking a short position in corn futures. The farmer's amount of corn and the timing of his/her harvest don't perfectly coincide with a standardized corn futures contract. The farmer would have to sell one or two corn futures contracts for either September or December, and would either be under hedged or over hedged.

Two General Types of Futures Contracts

1. Financial Futures : interest rate contracts (T-bonds, fed funds, Eurodollar, etc.) to manage interest rate risk, stock index contracts (SP500 Index, DJIA) to hedge stock price declines (portfolio insurance), currency contracts (DM, Yen, SF, etc.) to hedge ex-rate risk.
2. Commodity Futures : grains (corn, oats, soybeans, wheat, barley), metals (copper, gold, silver, platinum), livestock (hogs, cattle, pork bellies), foods and fibers (sugar, coffee, cotton, orange juice, rice), petroleum (crude oil, natural gas, heating oil, gasoline, propane), miscellaneous (lumber, seafood, electricity). These contracts allow participants to hedge against commodity price risk.

   Chicago Board of Trade started commodity futures trading in 1848 for agricultural products: grains, beef, pork bellies, etc. Chicago Mercantile Exchange started in 1874 as a rival exchange to CBT, and specialized originally in butter futures contracts.

   Now there are many futures exchanges: CBT, CME, NYM, CSCE (Coffee, Sugar and Cocoa Exchange), CTN (NY Cotton), TFE (Toronto Futures Exchange), MPLS (Maples Grain Exchange), and NYFE (NY Futures exchange), ME (Montreal Exchange), you can see these things in the WSJ.

   Futures contracts helped to stabilize volatile agricultural prices. Prices dropped sharply after harvest and then rose sharply when shortages developed later.

   About 2/3 of futures contracts are now for financial futures - bonds, stock indexes and currency and the other 1/3 is for agricultural commodities, metals and energy. Recent ranking of individual contracts by size (contracts traded): U.S. T-Bonds, S&P500 Stock Index, Eurodollars, Oil, Gold, Corn. Risk that markets are most worried about? Interest rate risk.

Futures Terms : BUY = GO LONG and SELL = GO SHORT. Example: In WSJ, July 2001 corn futures are trading at 238 cents per bushel, or $2.38/bushel. You can buy corn or sell corn at that price for delivery in July 2001, in units of 5,000 bushel per futures contract. If you buy a July 2001 corn futures contract, you are "going long" on corn. If you sell a July 2001 corn futures contract, you are "going short" on corn.

Proftit/Loss From Futures Contract

1. For the person SELLING CORN @ $2.38/bushel (short position), they will make a PROFIT when the spot/cash price of corn GOES BELOW $2.38/bushel and they will suffer a LOSS when the cash price GOES ABOVE $2.38/bushel.

   Reason: They have a contract to sell corn at $2.38/bu to the futures contract buyer (long position), and if corn goes to $2 in the cash market, they could theoretically buy low at the spot price ($2) and sell high at the contract price of $2.38 and make money. If corn goes to $3.00 bushel in the cash market, they would now have to buy high at $3/bu and sell low at $2.38, for a loss $0.62/bushel.

2. For the person BUYING CORN @ $2.38/bu (long position), they will make a PROFIT when the spot/cash price of corn GOES ABOVE $2.38/bu and they will suffer a LOSS when the cash price GOES BELOW $2.38/bushel.

   Reason: They have a contract to buy corn at $2.38/bu from the futures contract seller (short position), and if the spot/cash price goes to $3, they can buy low at $2.38 from the seller, and then sell high at $3 in the cash market and make money ($0.62/bu). However, if the cash/spot price goes to $2/bushel, they now have to buy high at $2.38 and sell low at $2.00, for a loss of $0.38/bushel.

   POINT: Futures markets are ZERO SUM trades, meaning that for every contract there is a winner and a loser and the winner wins the same amount as the loser loses, NET OUTCOME = 0 (+$1 winner, -$1 loser, ZERO SUM OUTCOME). For example, if spot prices for corn go up to $3/bushel, the long position makes $.62 profit and the short position loses $.62. If the cash price falls to $2, the short position makes $.32 profit per bushel and the long position has a loss of $.32 per bushel.

Two Types of Futures Markets Participants

1. **Hedgers** - futures traders who have a personal or business interest in the future commodity price, ex-rate or interest rate, e.g. importers/exporters, corporations buying and selling in the future, farmers, portfolio managers, firms expecting to borrow money in the future, firms/investors expecting to invest money in the future, etc.

Examples - farmers (sellers) and producers are worried about the price of their product going down in the future. They can use futures contract to lock in price now for future output of oil, corn, sugar, steel, gold, beef, pork, lumber, etc. by going SHORT on contracts for their product.

Buyers of commodities are worried about the prices of the products they buy going up in the future, they can protect against price risk by going LONG on commodities, e.g. GM going long on steel, Northwest Airlines going long on oil or gas contracts.

Exporters (importers) receiving foreign currency (paying in foreign currency) can hedge risk by going short (long) on currency futures.

A firm borrowing money in the future is worried about interest rates going up, bond prices going down, they would hedge interest rate risk by going short on TBond futures contracts. A firm investing money in the future is worried about interest rates going down, bond prices going up, they would hedge by going long on TBond futures.

Using Futures Contracts for Hedging Risk, i.e. "BUying Insurance"

To understand what position someone takes, always ask this question: **What is the party worried about?** What event do they want to insure against? What would represent an adverse price, currency or interest rate movement for that party? Once you identify what they are worried about, that determines the futures position they will take to protect against possible loss.

Price Risk

Buyer: worried about what? _____________ Futures Position: _____________

Seller: worried about what? _____________ Futures Position: _____________

Currency Risk

Exporter: receiving foreign currency, worried? _______ Futures_____________

Importer: paying in foreign currency, worried? ________ Futures_____________

Interest Rate Risk

Borrower: borrowing in the future, worried? ________ Futures _____________

Lender: lending money in future, worried? _________ Futures _____________

Derivative markets are actually insurance markets, and allows firms to manage, predict and control their revenue and expenses by locking in prices, interest rates and ex-rates ahead of time, to eliminate or minimize currency, price or interest rate risk. The future is uncertain, unpredictable and risky for businesses, and they can use futures contracts to hedge risk of future uncertainties.

2. **SPECULATORS** - have no personal or business interest in the commodity or currency, they are trading futures contracts as a purely speculative investment or gamble. For example, an investor could take a position on a corn futures contract for July 2001 @ $2.38/bushel, and they are not in the corn business, they have no interest in actually receiving or delivering corn at expiration, they are just taking a position on the price of corn in the future. Derivative concept, the corn futures contract "derives" its value from the price movements in an underlying commodity, e.g. corn, and actual delivery of corn never actually takes place. Speculators can participate in futures trading **because actual delivery is not required.**

   For example, if a speculator thinks the cash price of corn will go above $2.38/bushel sometime between now and July 2001, they take a LONG POSITION, and buy corn futures contracts. They are speculating that the P > $2.38, and will make money if that happens. They buy @$2.38/bu, and hope to "sell" at a price > $2.38/bushel, they make money at P > $2.38. Speculator is gambling (betting) that the price of corn will be > $2.38.

   If the speculator thinks the cash price of corn will go below $2.38/bushel, they take a SHORT POSITION, and sell corn futures. They will make money if the P < $2.38/bushel, they are betting that the price of corn will fall.

   What would be the advantages of having speculators in the futures markets, in addition to hedgers?

Example of Using Futures Contract to Hedge Price Risk

July 2001 corn is trading @ 238 in the WSJ (settle price), quoted in cents per bushel, so the futures price is currently $2.38/bu for July 2001 corn, and one contract is for 5,000 bushels of corn. The "open interest" is listed as 52,890, meaning there are almost 53,000 contracts outstanding right now for July 2001 corn. At the price of $2.38, you can either buy (go long) at that price or sell (go short) at that price. For each outstanding contract, there is a buyer (long) and a seller (short) who have agreed to buy or sell July 2001 corn. If you took a position today, you would either agree to buy or sell at $2.38/bushel.

Identifying the positions: Corn farmer, as a seller of corn, is worried about what? ________ What position do they take? ___________ Pepsi Cola, as a large corn buyer is worried about what? __________ What position do they take? _______________

Hedging Strategy: Corn seller is worried about corn prices falling, so they take a SHORT POSITION on corn, to put themselves in a position to make money on a futures contract if there is an adverse price movement (falling prices). Corn buyer is worried about corn prices rising, so they take a LONG POSITION on corn, to make money on a futures contract if the worst case scenario develops: rising corn prices.

Example: Suppose the corn farmer expects a crop of 100,000 bushels and is worried that the price of corn will fall below $2.38 by next July. Farmer protects his 100,000 bushel corn crop with 20 futures contracts by going short on corn futures and locks in a price of $2.38/bu. for next July.

Important Point: 98% of futures contract gets settled in cash, not the commodity, corn in this example! If the farmer had a forward contract for $2.38, he/she would make actual delivery of the corn to the buyer. But for a futures contract, the farmer DOES NOT deliver the corn, they settle the futures contract in cash, NOT corn. The farmer will therefore have two separate transactions at harvest:

1. **Cash transaction:** Sell 100,000 bushels of corn in July 2001 at the spot (cash) price at that time.
2. **Futures contract:** Settle the futures contract in cash in July 2001.

   The farmer uses the futures contract like an insurance contract to guarantee/lock-in a price of $2.38/bu and revenue of $238,000, but DOES NOT actually sell the corn for $2.38. The farmer **sells corn at the spot price** and uses the futures contract to guarantee a net price of $2.38 per bushel.

Example: a) Suppose the cash price for corn falls to $2.00/bu. by next July.

1. Farmer gets only $200,000 cash in the spot market from sale of crop, 100,000 bushels @ $2/bu. = $200,000 cash.
2. Farmer makes $38,000 profit on his/her short position in the corn futures contract. ($2.38 - $2.00) x 100,000 = $38,000. Logic: Farmer has a contract to sell 100,000 bu of corn at a price of $2.38 to the buyer (long), and they could theoretically now go out an buy it at the cash price of $2 and sell at $2.38, for a $0.38/bu profit.

Result

1. Cash Proceeds from the sale @ cash price of $2/bu = $200,000
2. <u>Cash Profit from short position on futures contract = $38,000</u>

FARMER'S NET REVENUE = $238,000 for 100,000 bushels

Therefore, the farmer nets the guaranteed price of $2.38 per bushel, $2 from the cash market, and 38 cents per bushel profit from the short position in the futures market.

b. Now suppose prices rise to $2.68/bu in the cash market by July 2001. Farmer now gets $268,000 cash from spot price for corn, but loses $30,000 from the short position on the futures contract. ($2.38 - 2.68) x 100,000 = -$30,000.

Result

1. Proceeds from sale @ cash price of $2.68/bu = $268,000
2. <u>Loss from short position on futures contract = ($30,000)</u>

FARMER'S NET REVENUE = $238,000 for 100,000 bushels.

Again, the farmer gets the guaranteed price of $2.38 per bushel, $2.68 in the cash market, minus the $0.30 loss on the futures contract.

POINT: Prices and go up or down, but the farmer has hedged price risk and locked in a price of $2.38 by going short on corn futures contracts. The farmer locks in a price of $2.38 and locks in total revenue of $238,000, and gives up the additional profit and revenue if corn actually goes up to $2.68, which would generate $268,000. However, the farmer is also protected against the worst case scenario of $2 per bushel and only $200,000 of revenue. Without the futures contract, the realized price for the farmer could range between $2 and $2.68/bushel, a 34% price spread; with the futures contract, the farmer gets a guaranteed, certain price of $2.38 per bushel. Farmer is in the farming business, not in the risk taking business.

Example of hedging for buyer: Royal Caribbean Cruise Lines uses futures contracts for June 2002 oil at $25/bbl. What are they worried about? _______ What position do they take? __________

Suppose spot prices go to $30/bbl by June 2002.

1. Gain on futures contract: ____________
2. Price in spot market: ________________

NET PRICE: __________________

Suppose spot prices go to $20/bbl by June 2002.

1. Loss on futures contract: _____________
2. Price in spot market: ________________

NET PRICE: __________________

Interest Rate Hedging Example

Example: Corporation is going to issue $10m worth of 15 year bonds in 60 days. Long term bond yields are currently at 10.75% and there is concern that rates will increase to 11% by the time the bonds are issued. The extra 1/4% would translate to an extra $25,000/yr in additional interest expense ($10m x .25%). The PV of the additional interest expense to the corporation over the next 15 years would be $179,772 as follows: n = 15, i = 11, PMT = $25,000, FV = 0, PV = ?

Worried about what???_____________________________________

Hedging Strategy to protect against interest rate risk: If you sell Tbond futures short for a 60 day maturity. If interest rates do rise and bond prices fall, you make money on the futures contract. If interest rates increase by .25%, you will make enough on the Tbond futures to cover the extra interest.

Example: S&Ls were and still are exposed to interest rate risk. Assets = Long-term mortgages, Liabilities = Short-term deposits. Value of S&L = PV Assets - PV Liab.
If interest rates go up, the PV of assets falls more than PV of liab. **Worried?:** Int. rates going up, bond prices going down. Position: go short on (sell) T-bond futures contracts to protect against int. rate risk. If interest rates go up, the Value of the bank fall, but the bank makes money on the futures contract to offset some or all of the loss. If int. rates fall, the value of the bank goes up, but there is a loss on the futures contract. In either case, the value of the bank is stabilized, protected from large fluctuations.

Cross-hedging - The corporate bond case above is an example of ***cross-hedging*** because Tbond futures are being used to hedge against interest risk for corporate bonds. We assume that interest rates on Tbonds and corporate bonds move together, but they may not always move perfectly together (e.g. the Treasury yield curve has been downward sloping recently) Perfect hedging is not always possible here because futures contracts are not traded for corporate bonds, only TBonds. Also, bond issue date may not match perfectly with future contract dates. (There are currently Tbond futures contracts for December, March and June only).

Partial Hedge : strategy where you only hedge part of the risk, e.g. hedge only $5m or $8m worth of bonds instead of the entire $10m.
Farmer: hedge only 50,000 bu instead of 100,000.

Using Futures Contracts To Hedge Currency Risk

Example:. U.S. exporter agrees to ship beef to UK in 6 months for a fixed amount of British pounds. Exporter will exchange pounds for dollars in 6 months. Worried?: British pound will depreciate. You can hedge and lock in a price today by selling British pound futures.

Example: Exporter agrees to sell 1 lb beef = 1 British pound. At current rates of $1.50/Br Pound, the US exporter would get $1.50/beef. If B Pound weakens and goes to $1.25/BP, the exporter will only get $1.25. It could strengthen and go to $1.75/BP, but that creates massive uncertainty. Assume a futures contract is available for $1.50/BP, exporter locks in at $1.50/lb. Worried? Pound falling. Goes short on British pound.

If worse case happens, and pound falls to$1.25/BP, the exporter loses .25 on beef income, but gains .25 on the futures contract, offsetting one another.

Importer: agrees to buy German wine in six months for a fixed amount of DMs, 10 DMs per bottle. Will take dollars and buy DMs in 6 months. At the current rate of $.65/DM, that would be $6.50/bottle. Worried about? The dollar getting weaker, the DM getting stronger. For example, if the ex-rate goes to $.75/DM, the cost would be $7.50/bottle, over a 15% increase. The importer would hedge by buying/going long on DM futures contracts.

Summary

Exporters: receiving foreign currency in future, worried about foreign currency getting weaker (dollar strengthening), sell currency futures.
Importers: paying in foreign currency in future, worried about dollar getting weaker, foreign currency getting stronger, buy currency futures.

Stock Index Futures Contracts : These are used to hedge against stock market declines, like buying "portfolio insurance." Contracts includes SP500 Index futures, DJIA futures, SP400 Midcap Index futures, Russell 2000, NASDAQ Index futures, etc. Hedging strategy for portfolio risk:

Mechanics of Futures Contracts

Hypothetical investment. It is May and we consider a Dec wheat contract at $4/bu. Contracts are for 5000 bu so the total contract is worth $20,000. We can control $20,000 of wheat with a small investment. Highly leveraged. Highly risky.

Margin requirement: the amount that has to be put up. Ranges from 2-10 percent depending on the contract. Wheat requires a $600 margin, or 3% of the total value of the contract. Much more highly leveraged than stock trading on margin - 50% requirement.

Margin maintenance requirements - like a minimum balance requirement. Usually 60-80% of the initial margin. For wheat, we assume that it is $400.

Daily settlement - accounts are settled daily to protect investors. If your account goes below $400, you need to put up additional money to cover the losses and get the margin account back to $600.

Assume that you take a long position. You buy wheat (go long) at $4, hoping that the price goes up.

A 4 cent reduction in wheat, from $4 to $3.96, would result in a loss of $200 (5000 x .04 = $200). Any additional movement would require additional margin funds. You would get a call from broker asking for additional margin funds. You could either close out your account or keep putting up money. 4 cents is a 1% movement, so you money can disappear very quickly.

A 3% movement would wipe out your entire investment (3%= $0.12 x 5000 = $600). On the other hand, assume that prices move in your favor and increase to $4.12/bu, a 3% increase (12 cents) within one month or even a few days. You make $600 (5000 bu x $0.12) for a return of 100% in one month:

($600 profit /$ 600 investment) x 100 = 100% yield (1200% annualized)

There is still five months to go on contract. You can cash out by reversing your position (going short cancels your long position), let the contract continue or double up and buy another one. The $600 gain is enough to buy another contract.

Price limits - To protect investors against losses and to minimize volatility, there are daily price limits on futures contracts. *Example*, corn and wheat can only move by 20 cents a bushel, up or down, before trading is discontinued for the day.

This is enough for today, I think.

APPROACHES TO SECURITY VALUATION

Objectives

- Understand valuation of risky securities
- Calculation of values of risky securities

As you all know that risk and returns are positively related, so it is better if You do some homework before investing your funds. This is where valuation of risky securities comes in to play.

The Valuation of Risky Securities

Payments received form reckless securities can be accurately predicted: Neither their amounts nor their timing is uncertain. But many securities do not meet such high standards. Some of all their payments are contingent on events with respect to amount, timing, or both. A bankrupt corporation may not make its promised bond payments in full or on time. A worker who is laid off may pay his or her bills late (or not at all). A corporation may reduce or eliminate its dividend if its business becomes unprofitable.

The security analyst must try to evaluate the circumstances affecting a risky investment's payments and enumerate the key events upon which such payments are contingent. For example, an aircraft manufacturer's fortunes may depend on whether or not the firm is awarded a major contract by the government, whether or not its recently introduced commercial aircraft is accepted by the airlines, or whether or not there is an upturn in the economy of such a company properly, the analyst must consider each of these contingencies and estimate the corresponding effect on the firm and its stock.

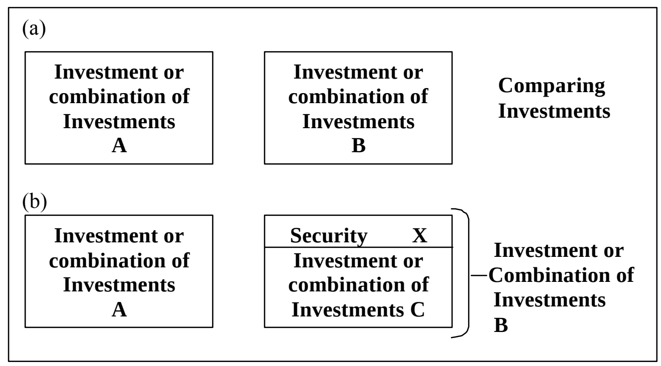
The identification of important influences and the evaluation to their impact are exceedingly difficult. Among other thins, the appropriate level of detail must be determined. The number of potentially relevant events is almost always very large, and the analyst must attempt to focus on the relatively few that appear to be most important. In some cases it may be best to differentiate only a few alternatives (for example, whether the economy will turn up, turn down, or stay the same). In some cases, finer distinctions may be needed (for example, whether the gross domestic product will be up 1%, 2 % or 3%).

Market Versus Personal Valuation

One approach to the valuation of risky securities focuses on the investor's personal attitudes and circumstances. Given his or her assessment of the likelihood of various contingencies, and feelings about the corresponding risks involved in an investment, an investor might determine the amount he or she would be willing to pay by some sort of introspection. This would be "personal" valuation of the security.

Such an approach would be appropriate if there were only one investment in the world, but such is not the case. A security need not and should not be valued without considering available alternatives. Current market values of other that nothing else is comparable. Security valuation should not be done in a vacuum; it should instead be performed in a market context.

Key to this approach is the comparison of one investment or combination of investments with other having comparable characteristics. For example, assume that A and B in the following figure (a) are the same in respect; then the two should be equal in value.



Now imagine that alternative B includes a security that an investor wishes to value – call it X. Moreover, assume that all other securities included in A and B are regularly traded and that their market values (prices) are widely reported and easily determined. Combination B can be thought to have two components: security X and the rest, which will be represented by C, as in the above figure (b) Combination C might include many securities, only one, or, as a very special case, none at all.

If people are willing to purchase combination A for $V_{A'}$ they should be willing to purchase combination B for the same amount, since the two provide comparable. Thus

$$V_A = V_{B'}$$

The value of B, however, will simply be the sum of the values of its components

$$V_B = V_X + V_C$$

This implies that the value of security X can be determined solely by reference to market values placed on the securities comprising combinations A and C. Since $V_A = V_{B'}$ it follows that

$$V_A = V_X + V_C$$

Or, by rearranging,

$$V_X = V_A - V_C$$

meaning that the value of X can be determined by subtracting the value of C from the value of A.

Approaches to Security Valuation

It is reasonable enough to say that market prices of "comparable investments" should be used to determine the value of a security. But when are two investments truly comparable?

An obvious case arises when investments provide identical payments in every possible contingency. If an investment's outcome is affected by relatively few events, it may be possible to purchase a set of other investments, each of which pays off in only one of the relevant contingencies. A properly selected mix of such investments may thus be completely comparable to the one to be valued. The next section illustrates this approach with an example drawn from the field of insurance.

A much more common approach to valuation is less detailed but more useful. Two alternatives are considered comparable if they offer similar expected returns and contribute equally to portfolio risk. Central to this view is the need to assess the probabilities of various contingencies.

Explicit Valuation of Contingent Payments

Insurance

Insurance policies are highly explicit examples of contingent payments. One can buy a $100,000 one-year "term" life insurance policy on a reasonably healthy 60-year-old for about $2,300. This of course, can be viewed as an investment (albeit a morbid one) : The sum of $100,000 will be paid by the insurance company if the insured dies within a year. Otherwise nothing at all will be paid. Involved is the sacrifice of a present certain value ($2,300) for a future uncertain value. The only relevant event is the possible death of the insured, and the relationship between that event and the amount to be paid is crystal clear.

Now imagine that a reasonably healthy 60-years-old executive asks you for a one-year loan. The executive would like as much as possible now; in return he or she promises to pay you $100,000 at the end of the year. Your problem is to determine the present value of that promise – that is, how much to advance now. But somewhat differently, you must determine an appropriate interest rate for the loan.

To keep the example simple, assume that the only source of uncertainty is the borrower's ability to remain in this position and thus earn the requisite money, and that this depends only on his or her continued presence among the living. In other words, if the borrower lives, the $100,000 will be repaid in full and on time; otherwise, you will receive nothing.

The piece of paper representing the executive's promise to pay $100,000 is your security X. What is the worth? The answer clearly depends significantly on the available alternatives. And a crucial factor is the current rate of interest.

Assume that the going rate for risk less one-year loans is 8%. If there were no doubt whatsoever that the executive would repay the loan, it would be reasonable to advance $92,592.59 (since $100,000/ $92,592.59 = 1.08). However, the uncertainty connected with the loan makes this inadvisable. The appropriate amount is obviously less. But how much less?

In this case an answer can easily be determined. It would be entirely reasonable to advance at least $90,292.59, making the "promised" interest rate on the

TABLE A

COST S AND PAYMENTS FOR A LOAN AND AN INSURANCE POLICY

Item	Event: Executive Dies	Event: Executive Lives	Cost
Loan	0	$100,000	$ 90,292.59
Insurance Policy	$ 100,000	0	$ 2,300.00
Total	$ 100,000	$100,000	$ 92,592.59

Figure A

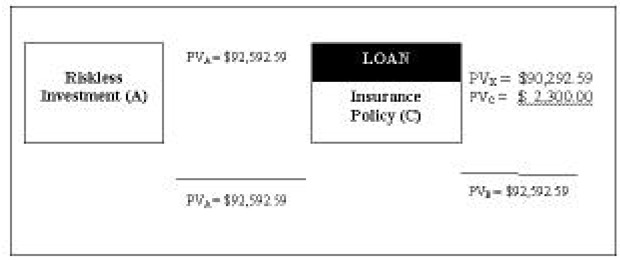


Figure A

Comparing Two Riskless Investments

Loan approximately 10.75% (since 100,000 / 90,292.59 = 1.1075). The basic for this calculation is quite simple. It relies on the fact that an investor can insure against the relevant risk, obtaining an overall position that is completely riskless.

Table A provides the details. The relevant event is whether or not the executive survives the year. The loan is thus risky investment, paying $100,000 only if the executive lives. The life insurance policy is also risky investment, paying $100,000 only if the executive dies. But a portfolio that includes both investments is totally riskless: Its owner will receive $100,000, no matter what happens! By paying $90,292.59 for the loan and $2,300 for the insurance policy, an investor could give up $92,592.59 now for a certain payment of $100,000 a year hence – obtaining a riskless return of 8%, which is the going rate on other riskless ventures.

This is of course, an application of the general procedure described in the previous section. Figure A summarizes the details in the format used earlier, for purposes of comparison.

Valuation in a Complete Market

Assume, for the present, that market values can be used to estimate the present value of any contingent payment. A market in which such detailed quotations are available is termed a **complete market**. While no real market confirms to this specification, it is useful to see how valuation would be done in such circumstances.

First, A way to represent the present value of a guaranteed commitment to pay $1 at a specified time if (and only if) a specified event or "state of the world" occurs is needed. The following will suffice:

$$PV(\$1, t, e)$$

Where

t= the time at which the dollar is to be paid,

e= the event that must occur if the dollar is to be paid.

Armed with this notation, any risky investment can now be analyzed. Every possible contingency could, in theory, be considered separately, giving a (probably very lengthy) list of contingent payments of the following form

Time of Payment	Event On Which Payment is Contingent	Amount of Payment
t_1	e_1	D_1
t_2	t_2	D_2
-	-	-
-	-	-
-	-	-

of course, some of the events might be the same, as might some of the times and amounts.

To find the present value of the investment, the present value of each of its contingent payments must be found and then added:

(1) Time of Payment	(2) Event on Which Payment is Contingent	(3) Amount of Payment	(4) Discount Factor	(5) = (3) X (4) Present Value
t_1	e_1	D_1	PV($1,$t_1$,$e_1$)	D_1XPV(($1,$t_1$,$e_1$)
t_2	e_2	D_2	PV($1,$t_2$,$e_2$)	D_2XPV(($1,$t_2$,$e_2$)
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
				Total value =__________

This **state-preference method** begins with the assumption that people's preferences are for *state-contingent claims* and concludes that securities will be valued on the basis of their payoffs in different "states of the world."

6.3.3 The limitations of Insurance

Some believe that Lloyd's of London will insure almost anything. Perhaps so. This could ease the security analyst's task considerably. He or she would only (!) have to determine the payments (D_1,D_2……..) associated with an investment, the times at which they cold be made ((t_1,t_2……) and the events on which they were contingent (e_1,e_2…..) The analyst could then use the premiums specified for the relevant insurance policies as estimates of appropriate discount factors [PV ($1, t_1, e1), (PV($1, t_2,e_2),….] , and perform the required calculations.

But even if Lloyd's will insure anything, the premiums charged for many policies might not attract any takers. There are a number of interrelated reasons for this. As a case in point, imagine an aerospace company, the future profits of which depends heavily on whether or not the firm will be awared a major government contract. Why not buy an appropriate insurance policy from Lloyd's, guaranteed to pay off if the firm loses the contract? Then only Lloyd's guaranteed to pay off if the firm loses the contract? Then only Lloyd's and the other firms in the industry would care about the outcome.

The idea is obviously whimsical. If Lloyd's were even willing to issue such a policy, the cost would be more than anyone would be likely to pay. Why? First, because of differences in information. Those familiar with the company or the government or both have better information about the likely outcome and can better assess the likelihood of various alternatives. Lloyd's operates at least partly in the dark. To protect itself, it will charge more than otherwise.

Second, there is the likelihood of adverse selection. If a policy of this sort is offered at a price low enough to attract anyone at all, the insurer can expect the firms that are least likely to win the contract to buy insurance, whereas those most likely to get the contract take their chances. This occurs frequently with life insurance. The less healthy an individual, the more likely he or she is to buy a policy. For this reason, the insured is usually required to pass a medical examination as a condition of sale. An examination of the health of a company's bid to win a government award might be much more difficult or expensive, so an insurance company must set its fees for such a policy on the assumption that it would end up insuring the riskiest client or clients.

Another factor is the thoroughly modern phenomenon described by the term **moral hazard**. The purchase of insurance may affect the likelihood of the event in question. If the manager of a firm is insured against the loss of the contract, he or she may well put less effort into the attempt to win it, increasing the likelihood of its loss and the insurance company's obligation to pay off. This explains more than its replacement value and the desire of many stockholders to have a corporation's officers own some of the firm's stock and none of its competitors' issues. Here again, the insurance company will account for this effect when setting prices.

Finally, there is the simple matter of overhead. Insurance people like to eat as do investors who provide the capital that insurance companies need. The costs of doing business will, over the long pull, be reflected in the prices charges for that business. No financial service is free, and insurance is no exception.

For all these reasons securities markets do not confirms to the specifications of the complete-market state preference model. Although the approach is helpful for addressing certain theoretical issues, it is less useful for investment purposes than the risk-return (or "mean-variance") approach, to which the discussion now turns.

PROBABILITY FORECASTING

Objectives

- Helps you in identifying various alternative outcomes and the probability.
- Understand the likelihood of various possible outcomes.
- Understanding of how to relate the expected values with the probabilities.

Dear friends, in the previous lesson you have come across of how the securities are valued and how risk is involved in it. And we discussed about the insurance policies and all other related factors. Now as a continuation of the process, lets start this lesson.

Lacking a plethora of widely available and low-cost insurance policies, it is not possible to value an investment without explicitly considering the likelihood of various outcomes. Instead, the analyst must attempt to assess directly the likelihood of each major event that can affect an investment. In short, he or she must engage in probabilistic forecasting.

The idea is simple enough, although its implementation is exceedingly difficult. The analyst expresses his or her assessment of the likelihood of every relevant event as a *probability.* If he or she feels that the chances of an event's taking place are 50-50, a probability of .50 is attached to the event. If the chances seem to be 3 out of 4, the probability is ¾ , or .75 (another way of expressing this is to say that the odds are 3 to 1 that the event will take place). If the analyst considers an event to be absolutely certain, a probability of 1.0 should be assigned. If he or she feels that an event is completely impossible, its probability of occurrence is zero.

It is important, of course, to be consistent in one's estimates. For example, if the events on a list are mutually exclusive and exhaustive (that is, one of them, but only one will take place), the probabilities should sum to 1.0.

Probability is, fundamentally, a subjective concept. Even simple cases fall under this heading. For example, a gambler may assess the portability of a coin's coming up heads at .5, based on knowledge of coins and past observations of the coin in question. But the estimate is still subjective, involving the implicit assumption that the coin really is "fair" and that the past is an appropriate guide to the future. Similar cases arise frequently in security analysis. Relative frequencies of various returns in the future. Clearly this procedure relies on assumptions that require subjective judgment and may in some circumstances be totally inappropriate. Forecasts based on the extrapolation of past relationships are neither wholly objective nor necessarily preferable to predictions obtained in more subtle ways.

Probabilistic forecasting entails a decision to confront uncertainty head-on, acknowledge its existence, and try to measure its extent. Instead of attempting to answer a question such as "What will General Motor earn next year?", the analyst explicitly considers some of the more likely alternatives and the likelihood of each one. This brings the analysis out in the open, allowing both the estimator and the users of such estimates to assess the reasonableness of the values. Insistence on a single number for each estimate, with no measure of associated uncertainty, would suggest naivete or insecurity on the part of the producer or the consumer of such predictions.

In some organizations, analysts engage in explicit probabilistic forecasting, passing on all their detailed assessments to others charged with bringing together the estimates made within the group. In other organizations, the analysts make explicit probabilistic forecasts but summarize their evaluations in a relatively few key estimates, sending only the latter to others. In still other organizations, analysts do not engage in explicit probabilistic forecasting. Instead, they produce estimates that summarize their implicit beliefs about the probabilities of various events. As always, it is not the form but the substance that matters.

It is often convenient to portray probabilistic forecasts graphically. The possible outcomes are represented on the horizontal axis and the associated probabilities on the vertical axis. Figure A provides an example. In this case the outcomes are qualitatively different in nature and can be listed only on the horizontal axis; the ordering and spacing are arbitrary.

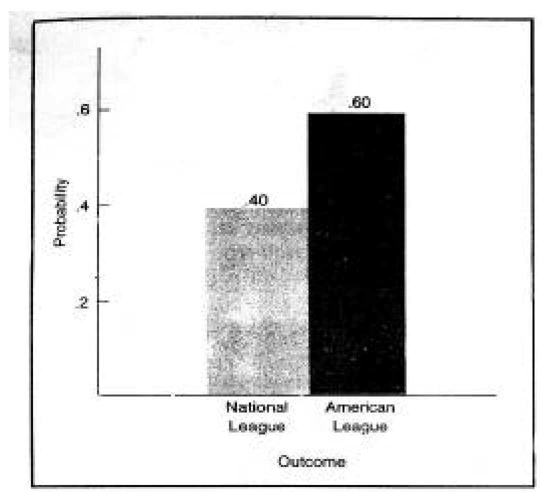


Figure B shows a somewhat different case. Here the alternative outcomes differ quantitatively and with regard to only one variable: earning per share next year. In this instance the analyst has chosen to group together all possibilities from $.90 to $.99, assess the portability that the actual amount will fall within that range, and then repeat the process for the range from $1.10 to $ 1.19 and other $.10 ranges.

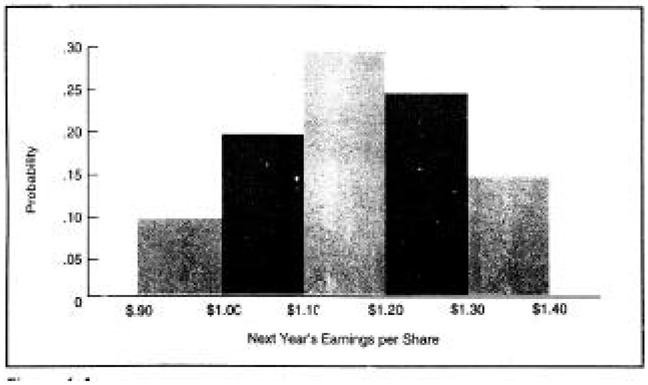


Figure B
Probabilities of Next Years Earnings per share

(Using Wide Ranges)

The analysis could, of course, have been conducted at a more detailed level, with probabilities estimated for outcomes in the ranges form $.90 to $.94, $.95 to $.99 and similar $.05 ranges. An even more detailed analysis would assign a probability to every possible outcome. In this case the bars would be numerous, and each would be very thin, as shown in Figure C. Note that the more numerous the number of bars, the smaller the sizes of associated probabilities.

The ultimate in a detailed prediction is represented by a continuous probability distribution. Such a curve represents, in effect, the tops of many thin bars. (Technically the curve represents what happens when there are an infinite number of bars.) Three examples of curves of this type are shown in figure D. Note that the vertical axis now measures probability density (instead of probability).

If continuous probability distributions are used, the analyst can forgo explicitly assessing particular individual outcomes. Instead, the analyst must draw a curve that seems to represent the situation as he or she sees it. The relative likelihood of any range of earnings is found by simply finding the size of the area under the curve but above the horizontal axis. Thus the likelihood of earnings being between $1.03 and $1.04 could be found by measuring the area under the curve between $1.03 and $1.04, which in this case is approximately .07 (that is, there is a chance of 7 out of 100 that earnings will be between $1.03 and $1.04 next year). With a discrete probability distribution such as those shown in figures 6.4 and 6.5, it was noted that the sum of the probabilities had to be 1.0. Now, with a continuous probability distribution, the total area under the curve must sum to 1.0.

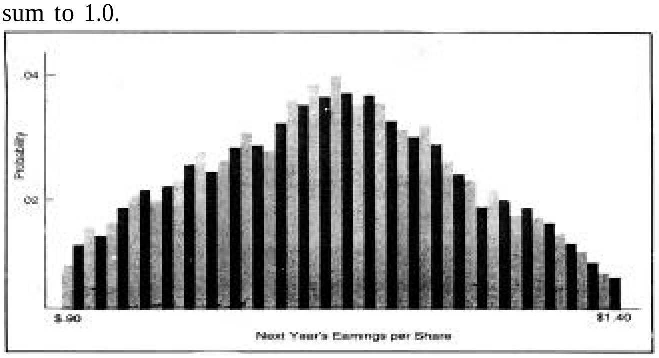


Figure C

Probabilities of Next Year's Earnings per Share (Using Narrow Ranges)

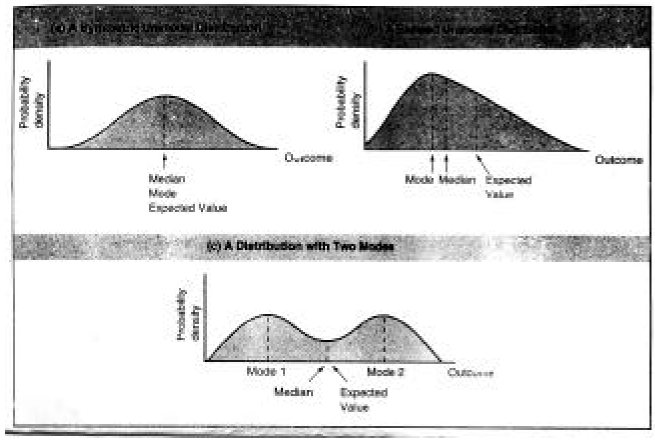


Figure D

Continuous Probability Distributions

Event Trees

When events follow one another over time or are in any sense dependent on one another, it is often useful to describe the alternative sequences with a tree diagram. Figure E provides an example.

A borrower has promised to pay $15 one year hence and $8 two years hence, if possible. The analyst feels that the odds are only 40 – 60 that the first payment will in fact be made in full. Otherwise, the analyst feels that the borrower will be able to pay only $10 one year hence.

As far as the second year is concerned , the likely situation depends, in this analyst's judgment, on the outcome in the first year. If the borrower manages to pay the full $ 15 in the first year, the analyst feels that the odds are only 1-9 that the borrower will be able to meet the $8 commitment at the end of two years. Otherwise, the borrower will pay less: $6. On the other hand, if the borrower pays out $10 in the first year, although there appears to be no chance of recovering the $5 shortfall, the analysts feels that the odds are about even (50-50) that the promised $8 will be paid in the second year. Of this does not happen, the analyst feels that $4 will be paid instead.

Figure 6.7 also shows the probability of each of the four possible sequences, or paths, through the tree. For example, the probability that both payments will be made in full is only .04, because there are only 40 chances out of 100 that the first payment will be made, and of those, only 1 out of 10 is expected to be followed by payment in full of the final obligation. This gives 4 out of 100 chances for the sequence: a probability of .04.

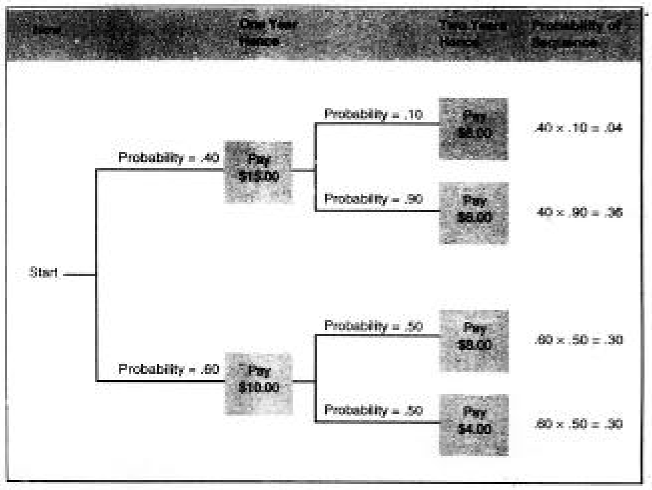


Figure E
An Event Tree

Expected Value

Often an analyst is uncertain about an outcome but wishes (or is required) to summarize the situation with one or two numbers – one indicating the *central tendency* of the distribution of outcomes and one measuring relevant risk. Both return and risk are discussed in subsequent chapters; the remainder of this chapter concentrates on the former.

How might a single number intended to summarize a set of possible outcomes be obtained? Obviously, no satisfactory way can be found if the alternative outcomes differ qualitatively (for example, the Nation League versus the American League in winning the World Series). But if the outcomes differ quantitatively, and especially if they differ in only one dimension, a number of possibilities present themselves.

Perhaps the most common procedure is to adopt the *most likely* value. This is known as the mode of the probability distribution (for a continuous probability distribution, the mode is the outcome with the highest probability density).

Another alternative is to provide a "50-50" number- a value that is as likely to be too low as it is to be too high. This is called the median of the probability distribution.

A third alternative is to use an expected value (also known as the mean), a weighted average of all the possible outcomes, using the associated probabilities as weights. In takes into account all the information expressed in the distribution, both the magnitude and the probability occurrence of each possible outcome. Almost any change in an investment's prospects or probabilities will affect the expected value of its outcomes (as it should).

In many instances there are no differences among these three measures. If the distribution is symmetric (each half is a mirror image of the other) and unimodal (there is most likely value), then the median, mode and expected value coincide. Thus an analyst may choose to think in terms of, say, a 50-50 (median) value, even though the number wanted is the expected value. Only if the underlying probability distribution is highly skewed might this procedure lead to difficulties.

In those cases in which the values do differ, there are good reasons to prefer the expected value case which the values do differ, there are good reasons to prefer the expected value. As stated earlier, it takes all the estimates into account. But it has another advantage. Estimates about the prospects for securities inputs for the process of portfolio is related in a straightforward way to the expected values of the returns for its securities, but neither the median nor the mode for portfolio can, in general, be determined from comparable values for its securities.

Table B provides an example of the computation of expected values. An analyst is trying to predict the impact on the prices of two securities of a surprise television address scheduled by the President. The analyst has delineated a number of possible announcements ranging from changes in the situation in the Middle East through a decision concerning the federal deficit. The alternatives represented in the table have been defined to be mutually exclusive and exhaustive (that is, every possible combination is shown in a different row). After much thought and with some trepidation, the analyst has also estimated the portability of each announcement and the resultant effect on the prices of the two securities. Finally, the analyst has computed the associated values of a portfolio containing one share of each stock.

TABLE B

ANALYSIS OF EFFECTS OF ANNOUNCEMENTS ON TWO SECURITIES AND A PORTFOLIO OF BOTH SECURITIES

Announcement	Probability	Predicted Price of Security A	Predicted Price of Security B	Predicted Value of A Portfolio of A and B
a	.10	$40.00	$ 62.00	$ 102.00
b	.20	42.00	65.00	107.00
c	.10	40.50	60.00	100.00
d	.25	41.00	61.00	102.00
e	.15	38.00	65.00	103.00
f	.10	40.50	59.00	99.50
g	.05	45.00	58.00	103.00
h	05	40.50	58.00	98.50
	Expected Values:	$40.73	$61.90	$ 102.63

The expected values are shown at the bottom of Table B. Each is obtained by multiplying the probability of every announcement by the associated price, then summing. For example, the expected price of security A is determined by computing [(.10 X $40.00) + (.20 X $42.00) + ….]; that of security B by computing [(.10 X $62.00) + (.20 X $65.00) + …}; and that of the portfolio by computing [(.10 X $102.00) + (.20 X $107.00)+….]. Not surprisingly, the expected value of the portfolio equals the sum of the expected values of its component securities. When the expected values for the securities are added together, one is, in effect, adding (.10 X $.40.00 + …) to (.10 X $62.00 + …). Clearly this will give the expected value of the portfolio, which is .10 X ($40.00 + $ 62.00) + ….

Expected Versus Promised Yield-to-Maturity

If payments from a bond are certain, there is no difference between the expected yield-to-maturity and the promised yield-to-maturity. However, may bonds fail to meet these standards? Two types of risk may be involved. First the issuer may defer some payments. A dollar received further in the future is, of course worth less in present value than a dollar received on schedule. Thus the present value of a bond will be smaller, the greater the likelihood of deferred payments. The second type of risk is potentially more serious. The borrower may default, in whole or in part, on some of the interest payments or on the principal at maturity. A firm becomes bankrupt when it is clearly unable to meet such obligations. The courts then divide the remaining assets among the various creditors in accordance with provisions agreed upon when the debts were issued.

To estimate the expected yield-to-maturity for a risky debt instrument, in principle all possible outcomes and the probability of each one should be considered. Assume that security in question cost $15; that is, the borrower wants $15 now in return for a commitment to pay $15 one year hence and $8 two years hence. The promised yield-to-maturity is the interest rate that makes the present value of these payments equal $15. In this case it is 38.15% per year, a substantial figure indeed.

But the analyst feels that the probability of actually receiving this yield-to-maturity is only .04. Table C shows the possible sequences (paths in the event tree), as well as the probability and the yield-to-maturity for each one. The expected yield-to-maturity is simply the weighted average of these values, using the probabilities as weights [for example, (.04X38.51%) + (.36 X 30.62%) + (.30 X 13.61%) + (.30 X -5.20%) = 15.09%].

The expected yield-to-maturity is considerably less than the promised amount: 15.09% as opposed to 38.51%. And the former is the more relevant figure for investment analysis. This is an important point. The yield-to-maturity is less than this figure; and the greater the risk, the greater the disparity. This is illustrated in Table D, which shows the (promised) yield-to-maturity values for six groups of industrial bonds classified by Standards & Poor's, a major rating service, as having different degrees of risk. Although the levels of all six yields reflect general interest rates at the time, the differences among them are primarily due to differences in risk. If the promised yields of all bonds were the same, the expected yields of high-risk bonds would be less than those of low-risk ones – an unlikely situation indeed. Instead, riskier bonds promise higher yields so that their expected yields can be at least as large as those of less risky ones.

The nature of most debt obligations would be more obvious if contracts were written somewhat differently. At present, a standard bond with no extra features "guarantees" that the borrower will pay the lender, say, $90 per year for 20 years, then $1,000 20 years hence.

Table C

PROMISED VERSUS EXPECTED YIELD-TO-MATURITY

Payment One Year Hence	Payment Two Years Hence	Probability	Yield-to-maturity
$15	$8	.04	38.51%
15	6	.36	30.62
10	8	.30	13.61
10	4	.30	- 5.20
		Expected yield-to-maturity=	15.09%

Table D

STANDARY & POOR'S INDUSTRIAL BOND YIELDS FOR AUGUST 1993

Rating	Yield-to-maturity
AAA	6.68%
AA	7.32
A	7.80
BBB	8.45
BB	9.11
B	10.57

xpected Holding-period Return

Calculating Holding-Period Return

Yield-to-maturity calculations do not take into account any changes in the market value of a security prior to maturity. This might be interpreted as implying that the owner has no interest in selling the instrument prior to maturity, no matter what happens to its price or his or her situation. The calculation also fails to treat intermediate payments in a fully satisfactory way. If the owner does not wish to spend interest payments, he or she might choose to buy more of these securities. But the number that can be bought at any time depends on the price at that time, and yield-to-maturity calculations fail to take this into account.

While few dispute the value of yield-to-maturity as an indicator of a bond's overall return, it should be recognized as no more than this. For some purposes other measures may prove more useful. Moreover, for other types of securities there is no maturity: Common stocks provide the most important example.

A measure that can be used for any investment is its **holding-period return.**

The idea is to specify a holding period of major interest, and then assume that any payments received during the period are reinvested. Although assumptions may differ from case to case, the usual procedure assumes the any payment received form a security (for example, a dividend from a stock, a coupon payment form a bond) is used to purchase more units of that security at the then current market price. Using this procedure, the performance of a security can be measured by comparing the value obtained in this manner at the end of the holding period with the value at the beginning. This **value- relative** can be concerted to a holding-period return by subtracting 1 form it: [1]

$$r_{hp} = \frac{\text{value at the end of the holding period}}{\text{value at the beginning of the holding period}} - 1$$

Holding-period return can be converted to an equivalent return per period. Allowing for the effect of compounding, the appropriate procedure is to find the value that satisfies the relationship

$$(1 + r_g)^N = 1 + r_{hp}$$

$$r_g = (1 + r_{hp})^{1/N}$$

where:

N = the number of periods in the holding period,

r_{hp} = the holding-period return,

r_g = the equivalent return per period, compounded every period.

Suppose that a stock sold for $46 per share at the beginning of one year, paid dividends of $1.50 during that year, sold for $50 at the end to the year, paid dividends of $2 during the next year, and sold for $56 at the end of that year. What was the return over the two-year holding period?

To simplify the calculations, assume that all dividend payments are received at year-end. Then the $1.50 received during the first year could have bought .03(= $1.50 / $50) shares of the stock at the end of the first year. In practice, of course, this would be feasible only if the money were pooled with other similarly invested funds—for example, in a mutual fund (the dividends form 100 shares could have been used to buy three additional shares). In any event, for each share originally held, the investor would have obtained $2.06 (=1.03 X $56) at the end of the second year. The ending value would thus have been $59.74(= $ 57.68 + $2.06), giving a value-relative of:

$$\frac{\$59.74}{\$46.00}$$

The holding-period return was thus 29.87% per two years. This is equivalent to $(1.2987)^{1/2} - 1 = .1396$, or 13.96% per year.

An alternative method of computation treats the overall value-relative as the product of value-relatives for the individual periods. For example, if V_o is the value at the beginning, V_1 the value at the end of the first year, and V_2 the value at the end of the second year:

$$\frac{V_2}{V_o} \quad \frac{V_2}{V_1} \quad \frac{V_1}{V_o}$$

Moreover, there is no need to carry the expansion in number of shares form period to period, as the factor (1.03 in the example) will simply cancel out in the subsequent periods' value-relatives. Each period can be analyzed in isolation, an appropriate value-relative calculated, and the set of such value-relatives multiplied together.

In our example, during the first year, ownership of a stock with an initial value of $46 led to stock and cash with a value of $1.50 at the end of the year. Thus,

$$\frac{V_1}{V_o} \qquad \frac{\$51.50}{\$46.00}$$

During the second year, ownership of stock with an initial value of $50 led to stock and cash with a value of $56 + $2 at year-end. Thus:

$$\frac{V_2}{V_1} \qquad \frac{\$58}{\$50}$$

The two-year holding period value-relative was therefore:

1.1196 X 1.16 = 1.2987

which is exactly equal to the value obtained earlier.

The value-relative for each period can be viewed as 1 plus the return for that period. Thus the return on the stock being analyzed was 11.96% in the first year and 16% in the second. The holding-period value-relative is the product of 1 plus each return. If N periods are involved.

$$\frac{V_N}{V_0}$$

To convert the result to a holding-period return stated as an amount per period with compounding, one can be find the *geometric mean return* of the periodic returns.

$$1 + r_g = [(1 + r_1)(1 + r_2)\ldots(1 + r_N)^{1/N}$$

More sophisticated calculations may be employed within this overall framework. Each dividend payment can be used to purchase shares immediately upon receipt, or alternatively, it can be allowed to earn interest in a saving account until the end of the period. Brokerage and other costs associated with reinvestment of dividends can also be taken into account, although the magnitude of such costs will undoubtedly depend on the overall size of the holding in question. The appropriate degree of complexity will, as always, be a function of the use for which the values are obtained.

Unfortunately, the most appropriate holding period is often at least as uncertain as the return over any given holding period. Neither an investor's situation nor his or her preferences can usually be predicted with certainty. Moreover, from a strategic viewpoint, an investment manager would like to hold a given security only as long as it outperforms available alternatives. Attempts to identify such periods in advance are seldom completely successful, but manager quite naturally continue to try to discover them. Holding-period return, like yield-to-maturity, provides a useful device for simplifying the complex reality of investment analysis. Although no panacea, it allows an analyst to focus on the most relevant horizon in given situation and offers a good measure of performance over such a period.

Estimating Expected Holding – Period Return

It is relatively straightforward matter to calculate a holding-period return after the fact. It is quite another thing to estimate it in advance. Any uncertainty surrounding payments by the issuer of a security during the period must be taken into account, but this usually much simpler than the task of estimating the end-of-period market values, which often constitute a large portion of overall return. For example, it might seem a simple matter to estimate the return over the next year for a share of Xerox stock. Dividends to be paid are often relatively easy to predict. But the price at year-end will depend on investors' attitudes toward the company and its stock at that time. To predict even a one-year holding period return one must consider a much longer period and assess not only the company's future but also investors' future attitudes about that future – a formidable task indeed.

Quite clearly, estimation of holding-period return must account in some way for uncertainty. If a single estimate is required, it should conform to the principles stated earlier. Explicitly, an expected value should be provided by considering both the various possibilities along with their probabilities. More specifically, a security's expected holding-period return is calculated as a weighted average of possible holding-period returns, using probabilities as weights.[2]

Expected Return and Security Valuation

There is very simple relationship between expected holding-period return, expected end-of period value, and current value:

$$\text{Expected holding-period return} = \frac{\text{expected end-of-period}}{\text{current value}} - 1$$

Thus

$$\text{Current value} = \frac{\text{expected end-of-period value}}{1 + \text{expected holding-period return}}$$

In words: To value a security, one needs to estimate the expected value at the end of a holding period and the expected return for the holding period that is appropriate for such a security.

The final phase is crucial. What is the "appropriate" expected return, and on what does it depends? Therein lies the remainder of the theory of valuation.

Questions and Problems

1. In March, a major bookmaker in Las Vegas accepted bets on the baseball teams that would eventually go to the World Series. For example, one could pay $10 at the time to bet that the Minnesota Twins would represent the American League in the World Series. The payoff on such a bet was set at $1,500 if the Twins did go to the World Series, and zero otherwise. Payoffs for bets on all teams in the America League Central Division were :

Team	Payoff per $1 Bet
Chicago White Sox	$180
Cleveland Indians	210
Kansas City Royals	60
Milwaukee Brewers	250
Minnesota Twins	150

a. What was the present vale of $1 contingent on the event (state of the world) "The Twins go to the World Series"?

b. What was the present value of $1 contingent on the event "The Brewers go to the World Series"?

c. Why did the answer for (a) and (b) differ?

d. If someone had offered to pay you $1 if any team in the American League Central Division went to the World Series, how much would you have paid for this bet ("security")? If you had been virtually certain that one of these teams would go to the World Series, would your answer differ? Why?

2. Mondovi Optical is a small business. Its owner, Tully Sparks, has requested that the local bank loan the firm $25,000 for two years. The federal government's Small Business Administration will fully guarantee such a loan for a $1,000 fee. If the riskfree two-year interest rate is 5% per annum, what is the interest rate that the bank should charge Mondovi?
3. Why is the insurance policy approach to risky security valuation so difficult to implement in practice?
4. From the perspective of an insurance company, provide two examples of adverse selection and two examples of moral hazard.
5. Distinguish between continuous and discrete probability distributions.
6. What are the advantages and disadvantages of using past investment results to assess the probabilities of future investment outcomes?
7. The average annualized return on the S&P 500 index of common stocks from 1926 through 1993 was 12.34%. If, on January 1, 1994, you had been required to provide an estimate of the expected return on the S&P 500 over the coming years, would you have chosen 12.34%? Why or Why not?
8. What is the value of event trees for the investment decision-making?
9. Consider Fort McCoy Company, whose stock currently sells for $10 per share. Dode Paskert, a financial analyst, has estimated the stock's potential year-end prices and associated probabilities over the next two years:

Year 1 The stock has a 30% chance of rising to $20. It has a 60% chance of rising to $12. It has a 10% chance of falling to $8.

Year 2 If the stock rises to $20 in year 1, it has a 50% chance of rising to $25 and a 50% chance of falling to $15. If the stock rises to $12 in year 1, it has a 70% chance of rising to $12.

a. Draw an event tree for Fort McCoy Company Stock.

b. Based on this event tree, calculate the stock's expected price at the end of year 2.

10. Calculate the expected return, mode and median for a stock having the following probability distribution.

Return	Probability of Occurrence
- 40%	.03
- 10	.07
0	.30
15	.10
30	.05
40	.20
50	.25

11 Bear Tracks Schmitz has estimated the following probability distribution of next year's dividend payments for Mauston

Inc's stock. According to Bear Tracks, what is the expected value of Mauston's dividend?

Dividend	Probability
$ 1.90	.05
95.95	.15
2.00	.30
2.05	.30
2.10	.15
2.15	.05

12. The probability distribution in Figure 6.6 (b) is "skewed to the right." Explain why the distribution's expected value is greater than the median, which in turn is greater than the mode.

13. Dupee Shaw is a fixed-income security analyst who is reviewing a bond issued by Wyeville Corporation. The bond has one year to maturity, at which time the company promises to pay $ 100. It currently sells for $90. Dupee believes that Wyeville may not pay full $100 at year-end. Dupee has estimated the following probability distribution of year-end payments. What is expected yield-to-maturity of the Wyeville bond according to Dupee?

Payment	Probability
$ 82	05
90	10
95	30
98	30
100	25

14. If an investment returns 7% per year, how long does it take for the investment's value to double?

15. Pol Perritt purchased 100 shares of Waunakee Inc. and held the stock for four years. Pol's holding-period returns over these four years were:

Year	Return
1	+20
2	+30
3	+50
4	-90

a. What was the value-relative of Pol's investment over the four-year period?

b. What was Pol's geometric mean return for the four-year period?

16. Stoughton Services stock currently sells for $40. it is expected to pay a dividend of $2 each year for the next several years. It just made its latest dividend payment. Pinky O' Neil expects to sell Stoughton stock two years from today at $50. The reinvestment rate is 5%. If this outcome occurs, what will be the equivalent compound annual return from holding Stoughton stock over this two-year period?

Distinguish between expected holding period return and yield-to-maturity

CAPITAL ASSETS PRICING MODEL & CAPITAL MARKET LINE

Objectives

- Understanding a specific set of assumptions about investor behaviour and existence of perfect security market.
- Understanding of Market portfolio which are weighted in proportion to its market values.
- Understand of how total risk of a security can be separated into market risk and non-market risk.

Dear friends, let's start this session. Hope you are all doing well with this subject.

An investor's optimal portfolio needs to estimate the expected returns and variances for all securities under consideration. Furthermore, all the co variances among these securities need to be estimated and risk free rate needs to be determined. Once this is done, the investor can identify the composition of the tangency portfolio as well as its expected return and standard deviation. At this juncture the investor can proceed to identify the optimal portfolio by noting where one of his or her indifference curves touches but does not intersect the efficient set. This portfolio involves an investment in the tangency portfolio along with a certain amount of either risk free borrowing or lending because the efficient set in linear (that is, a straight line.)

Such an approach to investing can be viewed as an exercise in **normative** economics, where investors are told what they should do. Thus, the approach is prescriptive in native. In this chapter the realm of **positive economics** is entered, where a descriptive model of how assets are prices is presented. The model assumes among other things that all investors use the approach to investing given in chapters, 7,8, and 9. The major implication of the model is the expected return of an asset will be related to a measure of risk for that asset known as *beta.* The exact manner in which expected return and beta related is specified by the **Capital Asset Pricing Model (CAPM)** basis for a number of the current practices in the investment industry. Although many of these practices are based on various extensions and modification of the **CAPM,** a sound understanding of the original version is necessary in order to understand them. Accordingly, this lesson presents the original version of the CAPM.

Assumptions

To see how assets are priced, a model (that is, theory) must be constructed. This requires simplification in that the model-builder must abstract from the full complexity of the situation and focus only on the most important elements. The way this is achieved is by making certain assumptions about the degree of abstraction that allows for some success in building the model. The reasonableness of the assumptions (or lack thereof) is of little concern. Instead the test of a model is its ability to help one understand and predict the process being modeled. As Milton Friedman, recipient of the 1976 Nobel Prize in Economics has stated in a famous essay.

[T]he relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic," for they never are, but whether they are sufficiently good approximately for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions.

Some of the assumptions behind the **CAPM** are also behind the normative approach to investing described in the previous three chapters. These assumptions are follows:-

1. Investors evaluate portfolios by looking at the expected returns and standard deviations of the portfolios over a one-period horizon.
2. Investors are never satiated, so when given a choice between tow otherwise identical portfolios, they will choose the one with the higher expected return.
3. Investors are risk-averse, so when given a choice between two otherwise identical portfolios, they will choose the one with the lower standard deviation.
4. Individual assets are infinitely divisible, meaning that an investor can buy a fraction of a share if he or she so desires.
5. There is a risk free rate at which an investor may either lend (that is, invest) money or borrow money.
6. Taxes and transaction costs are irrelevant.

To these assumptions the following ones are added :

7. All investors have the same one-period horizon.
8. The risk free rate have the same for all investors.
9. Information is freely and instantly available to all investors.
10. Investors have **homogeneous expectations**, meaning that they have the same perceptions in regard to the expected returns, standard deviations, and covariance's of securities.

As can be seen examining these assumptions, the **CAPM**, reduces the situation to an extreme case. Everyone has the same information and agrees about the future prospects for securities. Implicitly this means that investors analyze and process information in the same way. The markets for securities are **perfect markets**, meaning that there are no *frictions* to impede investing. Potential impediments such a finite divisibility, taxes, transaction costs, and different risk-free borrowing and lending rates have been assumed away. This allows the focus to change from how an individual should invest to what would happen to security prices if everyone invested in a similar manner. By examining the collective behavior of all investors in the market place, the nature of the resulting equilibrium relationship between each security's risk and return can be developed.

The Capital Market Line

The Separation Theorem

Having made these ten assumptions, the resulting implications can now be examined. First, investors would analyze securities and determine the composition of the tangency portfolio. In so doing, everyone would obtain in equilibrium the same tangency portfolio. However, this is not surprising because there is complete agreement among investors on the estimates of the securities' expected returns, variances, and covariances, as well as on the size of the riskfree rate. This also means that the linear efficient set (described in Chapter 9) is the same for all investors because it simply involves combinations of the agreed-upon tangency portfolio and either riskfree borrowing or leading.

As all investors face the same efficient set, the only reason they will choose different portfolios is that they have different indifferences curves. Thus different investors will choose different portfolios from the same efficient set because they have different preferences toward risk and return. For example, as was shown in figure. 9.8 , the investor in panel (a) will choose a different portfolio than the investor in panel (b). Note, however, that although the chosen portfolios will be different, *each investor will choose the same combination of risky securities,* denoted T in Figure 9.8. This means that each investor will spread his or her funds among risky securities in the same relative proportions, adding riskfree borrowing or lending in order to achieve a personally preferred overall combination of risk and return. This feature of the CAPM is often referred to as the separation theorem:

The optional combination of risky assets for an investors can be determined without any knowledge of the investor's preferences toward risk and returns.

In other words, the determination of the optimal combination of risky assets can be made without determining the shape of an investor's indifference curves.

The reasoning behind the separation theorem involves a property of the linear efficient set introduced in Chapter 9. There it was shown that all portfolios located on the linear efficient set involved an investment in a tangency portfolio combined with varying degrees of riskfree borrowing or lending. With a CAPM each person faces the same linear efficient set, meaning that each person will be investing in the same tengency portfolio (combined with a certain amount o either riskfree borrowing or lending that depends upon the person's indifference curves). It therefore follows that the risky portion of each person's portfolio will be the same.

Let's again consider three securities corresponding to the stock of Able, Baker and Charlie companies. With a risk free rate of return of 4%, the tangency portfolio *T* was shown to consist of investments in Able, Baker, and Charlie in proportions equal to .12, .19, and .69, respectively. If the ten assumptions of the CAPM, are made, then the investor shown in panel (a) of Figure 9.8 would invest approximately half of his or her money in the risk free asset and the remainder in *T*. The investor shown in panel (b), on the other hand, would borrow an amount equal to approximately half the value of his or her initial wealth and proceed to invest these borrowed funds as well as his or her own funds in T.[3] Thus the proportions invested in the three stocks for panel(a) and (b) investors would equal

Although the proportions to be invested in each of these three risky securities for the panel (a) investor (0.60, .095, .345) can be seen to be different in size from their values for the panel (b) investor (0.180, .285, 1.035), note how the relative proportions are the same, being equal to .12, .19, and .69, respectively.

$$(.5) \quad X \begin{bmatrix} .12 \\ .19 \\ .69 \end{bmatrix} = \begin{bmatrix} .060 \\ .095 \\ .345 \end{bmatrix} \quad \text{for the investor in part (a)}$$

$$(1.5) \quad X \begin{bmatrix} .12 \\ .19 \\ .69 \end{bmatrix} = \begin{bmatrix} .180 \\ .285 \\ 1.035 \end{bmatrix} \quad \text{for the investor in part (b)}$$

The Market Portfolio

Another important feature of the CAPM is that in equilibrium each security must have a nonzero proportion in the composition of the tangency portfolio. This mean that no security can in equilibrium have a proportion in T that is zero. The reasoning behind this feature lies in the previously mentioned separation theorem, where it was asserted that the risky portion of every investor's portfolio is independent of the investor's risk-return preferences. The justification for the theorem was that the risky portion of each investor's portfolio is simply an investment in T. If every investor is purchasing T and T does not involve an investment in each security, then nobody is investing in those securities with zero proportions in *T*. This means that the prices of these zero-proportion securities must fall, there by causing the expected returns of these securities to rise until the resulting tangency portfolio has a nonzero associated with them.

In the previous example, Charlie had a current price of $62 and an expected end-of-period price of $76.14. This meant that the expected return for Charlie was 22.8% [=($76.14 - $62)/$62]. Now imagine that the current price of Charlie is $72, not $62, meaning that its expected return is 5.8%[=($76.14 -$72)/$72]. If this were the case, the tangency portfolio associated with a riskfree rate of 4% would involve just Able and Baker in proportions of .90 and .10, respectively. Because Charlie has a proportion of zero, nobody would to hold shares of Charlie. Consequently, orders to sell would be received in substantial quantities with virtually no offsetting orders to buy being received. As a result, Charlie's price would fall as brokers would try to ding someone to buy the shares. However, as Charlie's price falls, its expected return would rise because the same end-of-period price of $76.14 would be forecast for Charlie as before and it would now cost less to buy one share. Eventually, as the price falls, investors would change their minds and want to buy shares Charlie. Ultimately, at a price of $62 investors will want to buy shares of Charlie so that in aggregate the number of shares demanded will equal the number of shares outstanding. Thus in equilibrium Charlie will have a nonzero proportion in the tangency portfolio.

Another interesting situation could also arise. What if each investor concludes that the tangency portfolio should involve a proportionate investment in the stock of Baker equal to .40, but at the current price of Baker there are not enough shares outstanding to meet the demand? In this situation orders to buy Baker will food in, and brokers will raise the price of in search of sellers. This will cause the expected return or Baker to fall, making it less attractive and thereby reducing its proportion in the tangency portfolio to a level where the number of shares demanded equals the number of shares outstanding.

Ultimately, everything will balance out. When all the price adjusting stops, the market will have been brought into equilibrium. First, each investor will want to hold, a certain positive amount of each risky security. Second, the current market price of each security will be at a level where the number of shares demanded equals the number of shares outstanding. Third, the riskfree rate will be at a level where the total amount of money borrowed equals the total amount of money lent. As a result, in equilibrium the proportions of the tangency portfolio will correspond to the proportions of what is known as the market portfolio, defined as follows:

The market portfolio is a portfolio consisting of all securities where the proportion invested in each security corresponds to its relative market value. The relative market value of a security is simply equal to the aggregate market value of the security divided by the sum of the aggregate market values of securities.

The reason the market portfolio plays a central role in the CAPM is that the efficient set consists of an investment in the market portfolio, coupled with a desired amount of either risk free borrowing or lending. Thus it is common practice to refer to the tangency portfolio as the market portfolio and to denote it as *M* instead of *T*. In theory, *M* consists not only of common stocks but also such other kinds of investments as bonds, preferred stocks, and real estate. However, in practice some people restrict *M* to just common stocks.

The Efficient Set

In the world of the Capital Asset Pricing Model it is simple matter to determine the relationship between risk and return for efficient portfolios. Figure 1 portrays it graphically. Point *M* represents the market portfolio and r_f represents the risk free rate of return. Efficient portfolios plot along the line starting at r_f and going through M and consist of alternative combinations of risk and return obtainable by combining the market portfolio with risk free borrowing or lending. This linear efficient set of the known as the Capital Market Line (CML). All portfolios other than those employing the market portfolio and risk free borrowing or lending would lie below the CML, although some light might plot very close to it.

The slope of the CML is equal to the difference between the expected return of the market portfolio and that of the risk free security ($r_m - r_f$) divided by the difference in their risks (s_m - 0) ($r_m - r_f$)/ s_m.[9] Because the vertical intercept of the CML is r_f, the straight line characterizing the CML has the following equation.

$$\bar{r}_p = r_f + \left[\frac{\bar{r}_M - r_f}{\sigma_m} \right] \sigma_p \qquad (1)$$

where r_p and s_p refer to the expected return and standard deviation of an efficient portfolio. In the previous example, the market portfolio associated with a risk free rate of 4% consisted of Able Baker, and Charlie (these stocks are assumed to be the only ones that exist) in the proportions of .12, .19, and .69, respectively. As it was shown in Chapter 9 that the expected return and standard deviation for a portfolio was 22.4% and 15.2, respectively, the equation for the resulting CML is:

$$\bar{r}_p = 4 + \left[\frac{22.4 - 4}{15.2} \right] \sigma_p = 4 + 1.21\, \sigma p$$

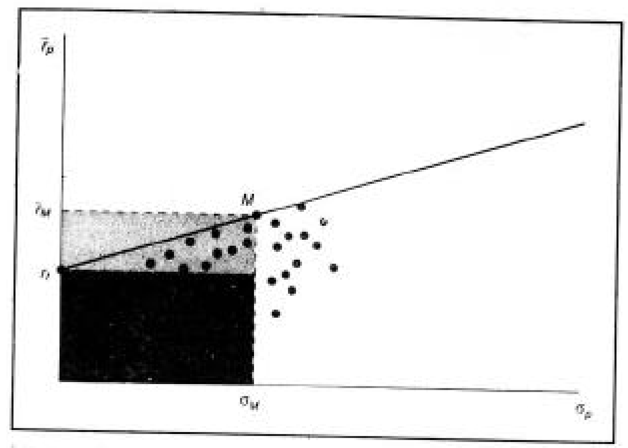


Figure 1

The Capital Market Line

Two key numbers can characterize equilibrium in the security market. The first is the vertical intercept of the CML (that is, the riskfree rate), which is often referred to as the *reward for waiting.* The second is the slope of the CML, which is often referred to as the *reward per unit of risk borne*. In essence, the security market provides a place where time and risk can be traded with their prices determined by the force of supply and demand. Thus the intercept and slope of the thought of as the price of the time and the price of risk, respectively. In the example, they are equal to 4% and 12.1, respectively.

Any way let's continue it in the next lesson as SML.

SECURITY MARKET LINE

Objectives

- After completion of this lesson you will be able to understand the relationships between co-variance and market.

By now CAPM must be clear to you, so that will help you while understanding this lesson: "THE SECURITY MARKET LINE (SML)". After completion of this lesson you will get better idea about CAPM.

Let's start with the implication of Individual Risky Assets

The Capital Market Line represents the equilibrium relationship between the expected return and standard deviation for the efficient portfolios. Individual risky securities will always plot below the line because a single risky security when held by it is an inefficient portfolio. The Capital Asset Pricing Model does not imply any particular relationship between the expected return and the standard deviation (that is, total risk) of an individual security. To say more about the expected return of an individual security, deeper analysis is necessary.

Following is the equation for calculating the standard deviation of any portfolio:

$$\sigma_P = \left[\sum_{i=1}^{N} \sum_{j=1}^{N} X_i X_j \sigma_{ij} \right]^{1/2} \qquad 1$$

where X_i and X_j denoted the proportions invested in securities *i* and *j*, respectively, and denoted the covariance of returns between security and i and j. Now consider using this equation to calculate the standard deviation of the market portfolio.

$$\sigma_M = \left[\sum_{i=}^{N} \sum_{j=1}^{N} X_{iM} X_{jM} \sigma_{ij} \right]^{1/2} \qquad 2$$

where X_{iM} and X_{jM} denote the proportions invested in securities *i* and *j* in forming the market portfolio, respectively. It can be shown that another way to write the above equation is as follows:

$$\sigma_M = \left[X_{1M} \sum_{j=1}^{N} X_{JM}\, \sigma_{1j} + X_{2M} \sum_{j=1}^{N} X_{JM\,2j} + X_{3M} \sum_{j=1}^{N} X_{jm}\, 3_j + \ldots\ldots\ldots\ldots + X_{NM} \sum_{j=1}^{N} X_{jm}\, \sigma_{Nj} \right]^{1/2} \qquad 3$$

At this point a property of covariance can be used: the covariance of security *i* with the market portfolio (σ_{im}) can be expressed as the weighted average of every security's covariance with security *i*:

$$\sum_{j=1}^{N} X_{JM}\, \sigma_{1j} = \sigma_{iM} \qquad 4$$

This property, when applied to each one of the *N* risky securities in the market portfolio, results in the following:

$$\sigma_M = [\, X_{1M}\, \sigma_{1M} + X_{2M}\, s_{2M} + X_{3M}\, \sigma_{3M} + \ldots\ldots + X_{NM}\, \sigma_{NM}]^{1/2} \qquad 5$$

Where s_{1M} denotes the covariance of security 1 with the market portfolio, s_{2M} denotes the covariance of security 2 with the market portfolio, and so on. Thus the standard deviation of the market portfolio is equal to square root of a weighted average of the covariance of all securities with it, where the weights are equal to the proportions of the respective securities in the market portfolio.

At this juncture an important point can be observed. Under the CAPM, each investor holds the market portfolio and is concerned with its standard deviation because this will influence the slope of the CML and hence the magnitude of his or her investment in the market portfolio. The contribution of each security to the standard deviation of the market portfolio can be seen Equation 5 to depend on the size of its covariance with the market portfolio. Accordingly each investor will note that the relevant measure of risk for a security is its covariance with the market portfolio, σ_{im} . This mean that securities with larger values of σ_{im} will be viewed by investors as contributing more to the risk of the market portfolio. It also means that securities with larger standard deviations should not be viewed as necessarily adding more risk to the market portfolio than those securities with smaller deviations.

From this analysis it follows that securities with larger values for have to provide proportionately larger expected returns to interest investors in purchasing them. To see why consider what would happen if such securities did not provide investors with proportionately larger levels of expected return. In this situation, these securities would contribute to the risk of the market portfolio. This means that deleting such securities from the market portfolio would cause the expected return of the market portfolio relative to its standard deviation to rise. Because investors would view this as a favorable change, the market portfolio would no longer be the optimal risky portfolio to hold. Thus security prices would be out of equilibrium.

The exact form of the equilibrium relationship between risk and return can be written as follows

$$\overline{r_i} = r_f + \left[\frac{\overline{r_m} - r_f}{\sigma^2_M} \right] \sigma_{iM} \qquad 6$$

As you can see in panel (a) of Figure A, Equation 6 represents a straight line having a vertical intercept of r_f and a slope of $[(r_m - r_f)\ s^2{}_m\ /]$. As the slope is positive, the equation indicates that securities with larger covariance with the market (s_{im}) will be prices so as to have larger expected returns (r_i). This relationship between covariance and expected return is known as the Security Market Line (SML).

Interestingly, a risky security with $s_{im} = 0$ will have an expected return equal to the rate on the riskfree security, r_f. Why? Because this risky security, just like the riskfree security, does not contribute to the risk of the market portfolio. This is so even though the risky has a positive standard deviation whereas the riskfree security deviation of zero.

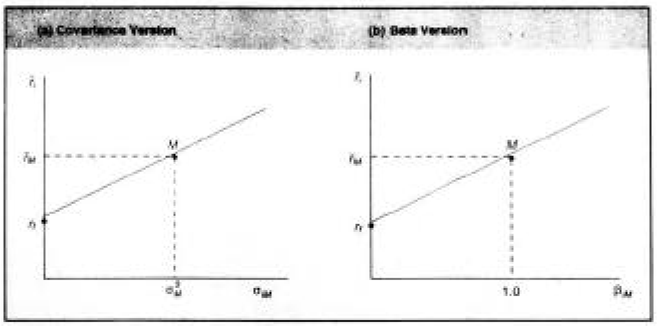


Figure A

The Security Market Line

It is even possible for some risky securities (meaning securities with positive standard deviations) to have expected returns less than the risk free rate. According to the CAPM, this will occur if $\sigma_{im} < 0$, thereby indicating that they contribute a negative amount of risk to the market portfolio to be lower than it would be if less money cause the risk of the market portfolio to be lower than it would be if less money were invested in them).

Also of interest is the observation that a risky security with $s_{im} = \sigma^2{}_{im}$ will have an expected return equal to the expected return on the market portfolio r_m, This is because such a security contributes an average amount of risk to the market portfolio.

Another way of expressing the SM is follows;

$$\overline{r}_i = r_f + (\overline{r}_M - r_f)\ b_{iM} \qquad 7$$

where the term is defined as:

$$b_{iM} = \frac{\sigma i_M}{\sigma^2_M} \qquad 8$$

The term is known as the beta coefficient (or simply the beta) for security *i*, and is an alternative way of representing the covariance of a security. Equation 7 is a different version of the SML as can be seen in panel (b) of Figure A. Although having the same intercept as the earlier version shown in Equation (6). r_f, it has a different slope. The slope of this version is $(r_m - r_f)$, whereas the slope of the earlier version was $[(r_m - r_f)/\ s^2{}_m]$.

One property of beta is that the beta of a portfolio is simply a weighted average of the betas of its component securities, where the proportions invested in the securities are the respective weights. That is, the beta of a portfolio can be calculated as:

$$\beta_{pM} = \sum_{i=1}^{N} X_i\ \beta_{iM} \qquad (9)$$

Earlier it was shown that the expected return of a portfolio is a weighted average of the expected returns of its component securities, where the proportions invested in the securities are the weights. This means that because every security plots on the SML, so will every portfolio. To put it more broadly, not only every security but also every portfolio must plot on an upward-sloping straight line in a diagram with expected return on the vertical axis and beta on the horizontal axis. This means that efficient portfolios plot on both the CML and the SML, although inefficient portfolios plot on the SML, but below the CML.

Also of interest is that the SML must go through the point representing the market portfolio itself. Its beta is 1 and its expected return is r_m, so its co-ordinates are $(1, r_m)$. Because riskfree securities have beta value of 0, the SML will also go through a point with an expected return of r_f and coordinates of $(0, r_f)$. This means that the SML will have a vertical intercept equal to r_f and a slope equal to the vertical distance between these two points$(r_m - r_f)$ divided by the horizontal distance between these two points $(1 – 0)$ or $(r_m - r_f)/(1 – 0) = (r_m - r_f)$. Thus these two points suffice to fix the location of the SML, indicating the appropriate expected returns for securities and portfolios with different beta values.

The equilibrium relationship shown by the SML comes to exist through the pressures on security prices. Given a set of security prices, investors calculate expected returns and covariance and then determine their optimal portfolios. If the number of shares of a security collectively desired differs from the number available, there will be upward or downward pressure on its price. Given a new set of prices, investors will reassess their desires for the variances securities. The process will continue until the number of shares collectively desired for each security the number available.

For the individual investor, security prices and prospects are fixed while the quantities are fixed (at least in the short run), and prices are variable. As in any competitive market, equilibrium requires the adjustment of each security's price until there is consistency between the quantity desired and the quantity available.

It may seem logical to examine historical returns on securities to determine whether or not securities have been priced in equilibrium as suggested by the CAPM. However, the issue of whether or not such testing of the CAPM can be done in a meaningful manner is controversial. For at least some purposes, affirmative test results may not be necessary to make practical use of the CAPM.

An Example is as Follows

If you refer the example in the last lesson, just try to recall, Able, Baker, and Charlie were shown to form the market portfolio in proportions equal to .12, .19 and .69 respectively. Given these proportions, the market portfolio was shown to have an expected return 22.4% and a standard deviation 15.2%. The risk free rate in the example was 4%. Thus for this example the SML, as indicated in Equation (6) is:

$$\bar{r}_i = r_f + \left[\frac{\bar{r}_m - r_f}{\sigma 2_m}\right] \sigma_{im} \qquad (6)$$

$$= 4 + \left[\frac{22.4 - 4}{(15.2)^2}\right] \sigma_{im}$$

$$= 4 + .08\ \sigma_{im} \qquad (10)$$

The following expected return vector and variance covariance matrix can be used in this example as

$$ER = \begin{bmatrix} 16.2 \\ 24.6 \\ 22.8 \end{bmatrix} \qquad VC = \begin{bmatrix} 146 & 187 & 145 \\ 187 & 854 & 104 \\ 145 & 104 & 289 \end{bmatrix}$$

At this point, the co variances of each security with the market portfolio can be calculated by using Equation (4). More specifically, the co variances with the market portfolio for Able, Baker, and Charlie are equal to:

$$\sigma_{1M} = \sum^{3} X_{JM}\, \sigma_{1j}$$

$$= (.12 \times 146) + (.19 \times 187) + (.69 \times 145)$$

$$= 153$$

$$\sigma_{2M} = \sum_{i=1}^{3} X_{JM}\, \sigma_{2j}$$

$$= (.12 \times 187) + (.19 \times 854) + (.69 \times 104)$$

$$= 257$$

$$\sigma_{3M} = \sum_{i=1}^{3} X_{JM}\, \sigma_{3j}$$

$$= (.12 \times 145) + (.19 \times 104) + (.69 \times 289)$$

$$= 236.$$

You just note how the SML as given in Equation (10) states that the expected return for Able should be equal to 4 + (0.08 x 153) = 16.2%. Similarly, the expected return for Baker should be 4 + (.08 x 257) = 24.6%, and the expected return for Charlie should be 4 + (.08 x 236) = 22.8%. Each one of these expected returns corresponds to the respective value given in the expected return vector.

Alternatively, Equation (8) can be used to calculate the betas for the three companies. More specifically, the betas for Able, Baker, and Charlie are equal to:

$$b_{1M} = \frac{\sigma_{1M}}{\sigma^2 m}$$

$$= \frac{153}{(15.2)^2}$$

$$= .66$$

$$b_{2M} = \frac{\sigma_{2M}}{\sigma^2 m}$$

$$= \frac{257}{(15.2)^2}$$

$$= 1.11$$

$$b_{3M} = \frac{\sigma_{3M}}{\sigma^2 m}$$

$$= \frac{236}{(15.2)^2}$$

$$= 1.02.$$

Now equation (7) indicated that the SML could be expressed in a form where the measure of risk for an asset was its beta. For the example under consideration, this reduces to:

$$r_i = r_f + (r_M - r_f)\, b_{iM}$$

$$= 4 + (22.4 - 4)\, b_{iM}$$

$$= 4 + 18.48\, b_{iM} \qquad (11)$$

Note how the SML as given in this equation states that the expected return for Able should be equal to 4 + (18.4 x .66) = 16.2%. Similarly, the expected return for Baker should be 4 + (18.4 x 1.11) = 24.6%, and the expected return for Charlie should be 4 + (18.4 x 1.02) = 22.8%. Each one of these expected returns correspond to the respective value given in the expected return vector.

It is important to realize that if any other portfolio is assumed to be the market portfolio, meaning that if any set of proportions other than .12, .19 and .69 is used, then such an equilibrium relationship between expected returns and betas (or covariance) will not hold. Consider a hypothetical market portfolio with equal proportions (that is, .333) invested in Able, Baker and Charlie. Because this portfolio has an expected return of 21.2% and a standard deviation of 15.5%, the hypothetical SML would be as follows:

$$\overline{r_i} = r_f + \frac{r_m - r_f}{\sigma^2_m} \sigma_{iM}$$

$$= 4 + \frac{21.2 - 4}{(15.5)^2} \sigma_{iM}$$

$$= 4 + .07\ \sigma_{iM}.$$

Able has a covariance with this portfolio of :

$$\sigma_{1m} = \sum_{j=1}^{3} X_{jM}\, \sigma_{1j}$$

$$= (.333 \text{ x } 146) + (.333 \text{ x } 187) + (.333 \text{ x } 145)$$

$$= 159,$$

which means that Able's expected return according to the hypothetical SML should be equal to 15.1% = 4 + (0.70 x 159). However because this does not correspond to the 16.2% figure that appears in the expected return vector, a portfolio with equal proportions invested in Able, Baker, and Charlie cannot be the market portfolio.

The Market Model

Market model assumes the return on a common stock was to be related to the return on a market index in the following manner.

$$r = \alpha_{iI} + \beta_{iI}\, r_I + \epsilon_{iI} \qquad (1)$$

Where r_i = return on security for some given period,

r_I = return on market index for the same period,

α_{iI} = intercept term,

β_{iI} = slope term,

ϵ_{iI} = random error term.

It is but natural to think about the relationship between the market model and the Capital Asset Pricing Model. After all, both models have a slope term called "beta" in them, and both models somehow involve the market. However, there are two significant differences between the models.

First, the market model is a factor model, or to be more specific, a single-factor model where the factor is a market index. Unlike the CAPM, however, it is not an *equilibrium model* that describes how prices are set for securities.

Second, the market model utilizes a market index such as the S$P 500, where as the CAPM involves the market portfolio. The market portfolio is a collection of all the securities in the marketplace, whereas a market index is in fact based on a sample of the market broadly construct (for example, 500 in the case of the S&P 500). Therefore, conceptually the beta of a stock based on the market model, b_{iI}, differs from the beta of the stock according to the CAPM, b_{iM}. This is because the market model beta is measured relative to a market index while the CAPM beta is measured relative to the market portfolio. In practice, however, the composition of the market portfolio is not precisely known, so a market index is used. Thus while conceptually different, betas determined with the use of a market index are treated as if they were determined with the use of the market portfolio. That is, b_{iI} is used as an estimate of b_{iM}.

In the example, only three securities were in existence – the common stocks of Able, Baker, and Charlie. Subsequent analysis indicated that the CAPM market portfolio consisted of these stocks in the proportions of .12, .19 and .69, respectively. It is against this portfolio that the betas of the securities should be measured. However, in practice they are like to be measured against a market index (for example, one that is based on just the stocks of Able and Charlie in proportions of .20 and .80 respectively).

Market Indices

One of the most widely known indices is the Standard & Poor's Stock Price Index (referred to earlier as the S&P 500), a value-weighted average price of 500 large stocks. Complete coverage of the stocks listed on the New York Stock Exchange is provided by the NYSE. Composite Index, which is broader than the S&P 500 in that it considers more stocks. The American Stock Exchange computes a similar index for the stocks it lists, and the National Association of Security Dealers provides an index of over-the-counter stocks traded on the Nasdaq system. The Russell 3000 and Wilshire 5000 stock indices are the most comprehensive indices of U.S. common stock prices published regularly in the United States. Because they consist of both listed and over-the-counter stocks, they are closer than the others to representing the overall performance of American stocks.

Without question the most widely quoted market index is the Dow Jones Industrial Average (DJIA). Although based on the performance of only 30 stocks and utilizing a less satisfactory averaging procedure, the DJIA provides at least a fair idea of what is happening to stock prices. Table 10.1 provides a listing of the 30 stocks whose prices have been reflected in the DJIA.

Market and Non-Market Risk

The total risk of a security could be partitioned into two components as follows –

$$\sigma^2_i = \beta^2_{i1}\, \sigma^2_1 + \sigma^2_{\varepsilon i} \qquad (2)$$

where the components are:

$\beta^2\, \sigma^2$ = market risk, and

σ^2 = unique risk.

Because beta, or covariance, is the relevant measure of risk for a security according to the CAPM, it is only appropriate to explore the relationship between it and the total risk of the security. It turns out that the relationship is identical to that given in Equation (2) except that the market portfolio is involved instead of a market index.

$$\sigma^2_{\mathrm{i}} = \beta^2_{iM}\, \sigma^2_{\mathrm{I}} + \sigma^2_{\varepsilon I} \qquad (3)$$

As with the market model, the total risk of security i, measured by its variance and denoted $s^2 i$, is shown to consist of two parts. The first component is the portion related to moves of the market portfolio. It is equal to the product of the square of the beta of the stock and the variance of the market portfolio, and also often referred to as the **market risk** of the security. The second component is the portion not related to moves of the market portfolio. It is denoted $s^2_{\hat{\imath} i}$ and can be considered *non-market risk*. Under the assumptions of the market model, it is unique to the security in question and hence is termed *unique risk*. Note that if is treated as an estimate of b_{im} , then the decomposition of s^2_i is the same in equation (2) and (3).

An Example

From the earlier example, the betas of Able, Baker, an Charlie were calculated to be .66, 1.11 and 1.02, respectively. As the standard deviation of the market portfolio was equal to 15.2%, this means that the material risk of the three firms is equal to $(.66^2 \times 15.2^2) = 100$, $(1.11^2 \times 15.2^2) = 285$, and $(1.02^2 \times 15.2^2) = 240$, respectively.

The non-market risk of any security can be calculated by solving equation (3) for $\sigma^2_{\in i}$

$$\sigma^2_{\in \mathrm{i}} = \sigma^2_{\mathrm{i}} - \beta^2_{\mathrm{iM}}\, \sigma^2_{\mathrm{M}} \qquad (4)$$

Thus, Equation (10.13) can be used to calculate the non-market risk of Able, Baker, and Charlie, respectively,

$$\sigma^2_{\in 1} = 146 - 100$$
$$= 46$$
$$\sigma^2_{\in 2} = 854 - 285$$
$$= 569$$
$$\sigma^2_{\in 3} = 289 - 240$$
$$= 49.$$

Non-market risk is sometimes expressed as a standard deviation. This is calculated by taking the square root of $s^2_{\hat{\imath} i}$ and would be equal to √46 = 6.8% for Able, Ö568= 23.9% for Baker, and √49=7% for Charlie.

Motivation for the Partitioning of Risk

At this point one may wonder: why partition total risk into two parts? For the investor, it would seem that risk is risk- whatever its source. The answer lies in the domain of expected returns.

Market risk is related to the risk of the market portfolios and to the beta of the security in question. Securities with larger betas will have larger amounts of market risk. In the world of the CAPM, securities with larger betas will have larger expected returns. These two relationships together imply that securities with larger market risks should have larger expected returns.

Non-market risk is not related to beta. This means that there is no reason why securities with larger amounts of non-market risks should have larger expected returns. Thus according to the CAPM, investors are awarded for bearing market but not bearing non-market risk. You try to solve the following problems and questions.

Questions and Problems

1. Describe the key assumptions underlying the CAPM.
2. Many of the underlying assumptions of the CAPM violated to some degree in the "real world". Does the fact invalidate the model's conclusions? Explain.
3. What is the separation theorem? What implications does it have for the optimal portfolio of risky assets held by investors?
4. What constitutes the market portfolio? What problems does one confront in specifying the composition of the true market portfolio? How have researchers and practitioners circumvented these problems?
5. In the equilibrium world of the CAPM, is it possible for a security not to be part of the market portfolio? Explain.
6. Describe the price adjustment process that equilibrates the market's supply and demand for securities. What conditions will prevail under such an equilibrium?
7. Will an investor who owns the market portfolio have to buy and sell units of the component securities every time the relative prices of those securities change? Why?
8. Given an expected return of 12% for the market portfolio, a riskfree rate of 6%, and a market portfolio standard deviation of 20%, draw the Capital Market Line.
9. Explain the significance of the Capital Market Line.
10. Assume that two securities constitute the market portfolio. Those securities have the following expected returns, standard deviations, and proportions:

Security	Expected Return	Standard Deviation	Proportion
A	10%	20%	40
B	15	28	60

    Based on this information, and given a correlation of .30 between the two securities and riskfree rate of 5%, specify the equation for the Capital Market Line.

11. Distinguish between the Capital Market Line and the Security Market Line.

12. The market portfolio is assumed to be composed of four securities. Their covariances with the market and their proportions are shown below:

Security	Covariance with market	Proportion
A	242	.20
B	360	.30
C	155	.20
D	210	.30

Given this data, calculate the market portfolio's standard deviation.

13. Explain the significance of the slope of the SML. How might the slope of the SML change over time?
14. Why should the expected return for a security be directly related to the security's covariance with the market portfolio?
15. The risk of a well-diversified portfolio to an investor is measured by the standard deviation of the portfolio's returns. Why shouldn't the risk of an individual security be calculated in the same manner?
16. A security with a high standard deviation of returns is not necessarily highly risky to an investor. Why might you suspect that securities with above-average standard deviations tend to have above-average betas?
17. Oil Smith, an investments student, argued, "A security with a positive standard deviation must have an expected return greater than the riskfree rate. Otherwise, why would anyone be willing to hold the security?" Based on the CAPM, is Oil's statement correct? Why?
18. Kitty Bransfield owns a portfolio composed of three securities. The betas of those securities and their proportions in Kitty's portfolio are shown on the next page. What is the beta of Kitty's portfolio?

Security	Beta	Proportion
A	.90	.30
B	1.30	.10
C	1.05	.60

19. Assume that the expected return on the market portfolio is 15% and its standard deviation is 21%. The riskfree rate is 7%. What is the standard deviation is 2%. The riskfree rate is 7%.What is the standard deviation of a well-diversified (no non-market-risk) portfolio with an expected return of 16.6%?
20. Given that the expected return on the market portfolio is 10%, the riskfree rate of return is 6%, the beta of stock A is .85, and the beta of stock B is 1.20:

a. Draw the SML.

b. What is the equation for the SML?

c. What are the equilibrium expected returns for stocks A and B?

d. Plot the two risky securities on the SML?

21. You are given the following information on two securities, the market portfolio, and the riskfree rate:

	Expected Return	Correlation with market Portfolio	Standard Deviation
Security 1	15.5%	0.90	20.0%
Security 2	9.2	0.80	9.0
Market Portfolio	12.0	1.00	12.0
Riskfree Rate	5.0	0.00	0.0

a. Draw the SML.

b. What are the betas of two securities?

c. Plot the two securities on the SML?

22. The SML describes an equilibrium relationship between risk and expected return. Would you consider a security that plotted above the SML to be an attractive investment? Why?
23. Assume that two securities, A and B, constitute the market portfolio. Their proportions and variances are .39, 160, and .61, 340, respectively. The covariance of the two securities is 190. Calculate the betas of two securities.
24. The CAPM permits the standard deviation of a security to be segmented into market and non-market risk. Distinguish between the two types of risk.
25. Is an investor who owns any portfolio of risky assets other than the market portfolio exposed to some non-market risk? Explain.
26. Based on the risk and return relationship of the CAPM, supply values for the seven missing data in the following table.

Security	**Expected Return**	**Beta**	**Standard Deviation**	**Non-market Risk ($\sigma^2_{\epsilon i}$)**
A	______%	0.8	_______%	81
B	19.0	1.5	_______	36
C	15.0	____	12	0
D	7.0	0	8	_______
E	16.6	_____	15	_______

27. (Appendix Question) Describe how the SML is altered when the riskfree borrowing rate exceeds the riskfree lending rate.

LESSON 16:
TUTORIAL

ARTICLE ON OPTION PRICING

OptionPricing: A Simplified Approach

Introduction

An option is a security which gives its owner the right to trade in a fixed number of shares of a specified common stock at a fixed price at any time on or before a given date. The act of making this transaction is referred to as exercising the option. The fixed price is termed the striking price, and the given date, the expiration date. A call option gives the right to buy the shares; a put option gives the right to sell the shares.

Options have been traded for centuries, but they remained relatively obscure financial instruments until the introduction of a listed options exchange in 1973. Since then, options trading has enjoyed an expansion unprecedented in American securities markets.

Option pricing theory has a long and illustrious history, but it also underwent a revolutionary change in 1973. At that time, Fischer Black and

Our best thanks go to William Sharpe, who first suggested to us the advantages of the discrete-time approach to option pricing developed here. We are also grateful to our students over the past several years. Their favorable reactions to this way of presenting things encouraged us to write this article. We have received support from the National Science Foundation under Grants Nos. SOC-77-18087 and SOC-77-22301.

Myron Scholes presented the first completely satisfactory equilibrium option pricing model. In the same year, Robert Merton extended their model in several important ways. These path-.breaking articles have formed the basis for many subsequent academic studies.

As these studies have shown, option pricing theory is relevant to almost every area of finance. For example, virtually all corporate securities can be interpreted as portfolios of puts and calls on the assets of the firm. [1] Indeed, the theory applies to a very general class of economic problems - the valuation of contracts where the outcome to each party depends on a quantifiable uncertain future event.

Unfortunately, the mathematical tools employed in the Black-Scholes and Merton articles are quite advanced and have tended to obscure the underly-ing economics. However, thanks to a suggestion by William Sharpe, it is possible to derive the same results using only elementary mathematics.[2]

In this article we will present a simple discrete-time option pricing formula. The fundamental economic principles of option valuation by arbitrage methods are particularly clear in this setting. Sections 2 and 3 illustrate and develop this model for a call option on a stock which pays no dividends. Section 4 shows exactly how the model can be used to lock in pure arbitrage profits if the market price of an option differs from the value given by the model, In section 5, we will show that our approach includes the Black-s school model as a special limiting case. By taking the limits in a different way, we will also obtain the Cox-Ross (1975) jump process model as another special case,

Other more general option pricing problems often seem immune to reduction to a simple formula. Instead, numerical procedures must be employed to value these more complex options. Michael Brennan and Eduardo Schwartz (1977) have provided many interesting results along these fines. However, their techniques are rather complicated and are not directly related to the economic structure *of* the problem. Our formulation, by its very construction, leads to art alternative numerical procedure which is both simpler, and for many purposes, computationally more efficient.

Section 6 introduces these numerical procedures and extends the model to include puts and calls on stocks which pay dividends. Section 7 concludes the paper by showing how the model can be generalized in other important ways and discussing its essential role in valuation by arbitrage methods.

[1] To take an elementary case, consider a firm with a single liability of a homogeneous class of pure discount bonds. The stockholders then have a 'call' on the assets of the firm which they can choose to exercise at the maturity date of the debt by paying its principal to the bondholders. In turn, the bonds can be interpreted as a portfolio containing a default-free loan with the same face value u the bonds and a short position in a put on the assets of the firm. .

[2] Sharpe (1918) has partially developed this approach to option pricing in his excellent new book, Investments. Rendleman and Bartter (1978) have recently independently discovered a similar formulation of the option pricing problem.

The Basic Idea

Suppose the current price of a stock is S = $50, and at the end of a period of time. its price must be either S* = $25 or S* = $100. A call on the stock is available with a striking price of *K* =$50, expiring at the end of the period.[3] It is also possible to borrow and lend at a 25 % rate of interest. The one piece of information left unfurnished is the current value of the call, C. However. if risk less profitable arbitrage is not possible, we can deduce from the given information *alone* what the value of the call *must* be!

Consider forming the following levered hedge:

1. *Write 3 calls* C *each,*
2. *Buy 2 shares at $50 each, and*
3. *Borrow $40 at 25%, to be paid back at the end of the period.*

Table 1 gives the return from this hedge for each possible level of the stock price at expiration. Regardless of the outcome, the hedge exactly breaks even on the expiration date. Therefore, to prevent profitable risk less arbitrage, its current cost must be zero; that is,

3C -100+40=0.

The current value of the call must then be C = $20.

Table I

Arbitrage table illustrating the formation of a risk less hedge.

	Present Date	Expiration Date	
		$S^* = \$25$	$S^*=100$
Write 3 calls	3C	-	-150
Buy 2 shares	-100	50	200
Borrow	40	-50	-50
Total		-	-

If the call were not priced at $20, a sure profit would be possible. In particular, if C = $25, the above hedge would yield a current cash inflow of $15 and would experience no further gain or loss in the future. On the other hand, if C = $15, then the same thing could be accomplished by buying 3 calls selling short 2 shares, and lending $40.

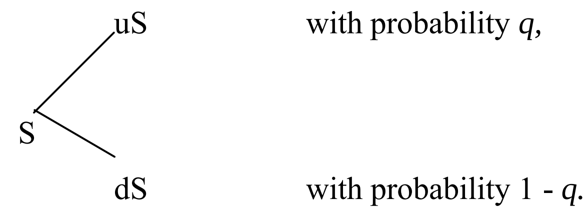
[3]To keep matters simple, assume for now that the stock will pay no cash dividends during the life of the call. We also ignore transaction costs margin requirements and taxes.

Table 1 can be interpreted as demonstrating that an appropriately levered position in stock will replicate the future returns of a call. That is, if we buy shares and borrow against them in the right proportion, we can, in effect, duplicate a pure position in calls. In view of this, it should seem less surprising that all we needed to determine the exact value of the call was its striking price, underlying stock price, range of movement in the underlying stock price, and the rate of interest. What may seem more incredible is what we do not need to know: among other things, we do not need to know the probability that the stock price will rise or fall. Bulls and bears must agree on the value of the call, relative to its underlying stock price!

This example is very simple, but it shows several essential features of option pricing. And we will soon see that it is not as unrealistic as its seems.

The Binomial Option Pricing Formula

In this section, we will develop the framework illustrated in the example into a complete valuation method. We begin by assuming that the stock price follows a multiplicative binomial process over discrete periods. The rate of return on the stock over each period can have two possible values: $u - 1$ with probability q, or d - 1 with probability 1 - q. Thus, if the current stock price is S, the stock price at the end of the period will be either uS or dS. We can represent this movement with the following diagram:



We also assume that the interest rate is constant. Individuals may borrow or lend as much as they wish at this rate. To focus on the basic issues, we will continue to assume that there are no taxes, transaction costs, or margin requirements. Hence, individuals are allowed to sell short any security and receive full use of the proceeds.[4]

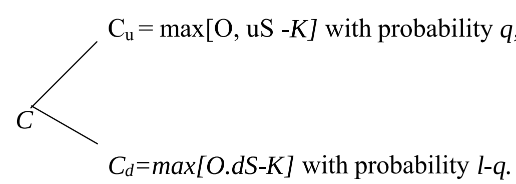
Letting r denote one plus the risk less interest rate over one period, we require $u > r > d$. If these inequalities did not hold, there would be profitable risk less arbitrage opportunities involving only the stock and riskless borrow-ing and lending.[5]

To see how to value a call on this stock, we start with the simplest situation: the expiration date is just one period away. Let C be the current value of the call, C_u be its value at the end of the period if the stock price

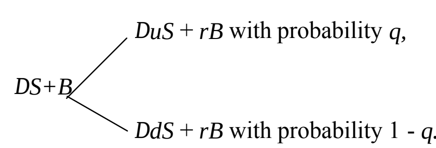
[4]Of course, restitution is required for payouts made to securities held short.

[5]We will ignore the uninteresting special case where q is zero or one and $u=d=r$.

goes to uS. and C_d be its value at the end of the period if the stock price goes to dS. Since there is now only one period remaining in the life of the call. we know that the terms of its contract and a rational exercise policy imply that Cu=max[0,dS-K] and Cd=max[0,dS-K]. Therefore.



Suppose we form a portfolio containing • shares of stock and the dollar amount B in risk less bonds.6 This will cost • S + B. At the end of the period, the value or this portfolio will be



Since we can select Δ and B in any way we wish, suppose we choose them to equate the end-or-period values of the portfolio and the call for each possible outcome. This requires that

$$DuS + rB = C_u$$

$$DdS + rB = C_d$$

Solving these equations. we find

$$D=\frac{C_u - C_d}{(u-d)S} \qquad B = \frac{uC_d - dC_u}{(u-d)r} \qquad (1)$$

With Δ and *B* chosen in this way, we will call this the hedging portfolio.

If there are to be no risk less arbitrage opportunities, the current value of the call, C. cannot be less than the current value of the hedging portfolio, *D* S + *B.* If it were, we could make a riskless profit with no net investment by buying the call and selling the portfolio. It is tempting to say that it also cannot be worth more. since then we would have a risk less arbitrage opportunity by reversing our procedure and selling the call and buying the portfolio. But this overlooks the fact that the person who bought the call we sold has the right to exercise it immediately.

"Buying bonds is the same as lending; selling them is the same' as borrowing.

Suppose that *DS+B<S-K.* If we try to make an arbitrage profit by selling calls for more than *DS + B,* but less than S - *K,* then we will soon find that we are the source of arbitrage profits rather than their recipient. Anyone could make an arbitrage profit by buying our calls and exercising them immediately.

We might hope that we will be spared this embarrassment because everyone will somehow find it advantageous to hold the calls for one more period as an investment rather than take a quick profit by exercising them immediately. But each person will reason in the following way. If I do not exercise now, I will receive the same payoff as a portfolio with *D*S in stock and *B* in bonds. If I do exercise now, I can take the proceeds, S -*K,* buy this same portfolio and some extra bonds as well, and have a higher payoff in every possible circumstance. Consequently, no one would be willing to hold the calls for one more period.

Summing up all of this, we conclude that if there are to be no riskless arbitrage opportunities, it must be true that

$$C=DS+B$$

$$= \frac{C_u - C_d}{u-d} + \frac{C_d - dC_u}{(u-d)r}$$

$$= \left[\frac{(r-d)}{(u-d)} C_u + \frac{(u-r)}{(u-d)} C_d \right] \Big/ r, \qquad (2)$$

if this value is greater than S - *K,* and if not, C = S - *K.* [7]

Eq. (2) can be simplified by defining

$$p? \frac{r-d}{u-d} \quad and \quad 1-p ? \frac{u-r}{u-d}$$

So that we can write

$$C = [pCu + (1- p)Cd]/r. \qquad (3)$$

It is easy to see that in the present case, with no dividends, this will always be greater than *S – K* as long as the interest rate is positive. To avoid

[7]1n some applications of the theory to other areas, it is useful to consider options which can be exercised only on the expiration date. These are usually termed European options. Those which can be exercised at any earlier time as well, such as we have been examining here, are then referred to as American options. Our discussion could be easily modified to include European calls. Since immediate exercise is then precluded, their values would always be given by (2), even if this is less than S -*K.*

spending time on the unimportant situations where the interest rate is less than or equal to zero, we will now assume that r is always greater than one. Hence, (3) is the exact formula for the value of a call one period prior to expiration in terms of S, *K,* u, *d,* and *r.*

To confirm this note that if *uS* • *K,* then *S* < *K* and C = 0, so *C > S – K.* Also if *dS* • *K,* then *C=S - (K/r) > S – K.* The r emaining possibility is *uS>K>dS.* In this case, *C = p(uS – K) / r.* This is greater than *S – K* if (1 – p) *dS > (p-r) K,* which is certainly true as long as *r>1.*

This formula has a number of notable features. First, the probability *q* does not appear in the formula. This means, surprisingly, that even if different investors have different subjective probabilities about an upward or downward movement in the stock, they could still agree on the relationship of C to S. u. *d,* and *r.*

Second, the value of the call does not depend on investors' attitudes toward risk. In constructing the formula, the only assumption we made about an individual's behavior was that he prefers more wealth to less wealth and therefore has an incentive to take advantage of profitable riskless arbitrage opportunities. We would obtain the same formula whether investors are risk-averse or risk-preferring.

Third, the only random variable on which the call value depends is the stock price itself. In particular, it does not depend on the random prices of other securities or portfolios, such as the market portfolio containing all securities in the economy. If another pricing formula involving other variables was submitted as giving equilibrium market prices, we could immediately show that it was incorrect by using our formula to make risk less arbitrage profits while trading at those prices.

It is easier to understand these features .if it is' remembered that the formula is only a relative pricing relationship giving C in terms of S, U, *d,* and *r.* Investors' attitudes toward risk and the characteristics of other assets may indeed influence call values indirectly, through their effect on these variables, but they will not" be separate determinants of call value.

Finally, observe that *p* a"*(r – d) / (u – d)* is always greater than zero and less than one, so it has the properties of a probability. In fact, *p* is the value *q* would have in equilibrium if investors were risk-neutral. To see this, note that the expected rate of return on the stock would then be the riskless interest rate, so

q(uS)+ (1 – q) (dS)=rS,

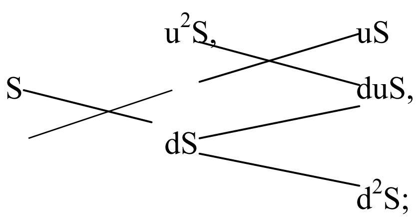
and

q = (r – d) / (u – d)=p.

Hence, the value of the call can be interpreted as the expectation of its discounted future value in a risk-neutral world. In light of our earlier observations, this is not surprising. Since the formula does not 'involve *q* or any measure of attitudes toward risk, then it must be the same for any set of, preferences, including risk neutrality.

It is important to note that this does not imply that the equilibrium expected rate of return on the call is the risk less interest rate. Indeed, our argument has shown that, in equilibrium, holding the call over the period is exactly equivalent to holding the hedging portfolio. Consequently, the risk and expected rate of return of the call must be the same as that of the hedging portfolio. It can be shown that Δ • 0 and B • 0, so the hedging portfolio is equivalent to a particular levered long position in the stock. In equilibrium, the same is true for the call. Of course, if the call is currently mispriced, its risk and expected return over the period will differ from that of the hedging portfolio.

Now we can consider the next simplest situation: a call with two periods remaining before its expiration date. In keeping with the binomial process, the stock can take on three possible values after two periods,



similarly, for the call,

C_{uu} = max [0, u^2 S – K]

C_u

C $\quad$ C_{du} = max [0, duS – K],

C_d

C_{dd} = max [0, d^2 S – K]

C_{uu} stands for the value of a call two periods from the -current time if the stock price moves upward each period; C_{du} and C_{dd} have analogous definitions.

At the end of the current period there will be one period left in the life of the call and we will be faced with a problem identical to the one we just solved. Thus, from our previous analysis, we know that when there are two periods left,

$C_u=[pC_{uu}+(1-p)C_{ud}] / r,$

and

$C_d = [pC_{du} + (1 - P)C_{dd}] / r.$ (4)

Again we can select a portfolio with DS in stock and B in bonds whose end-of-period value will be C.. if the stock price goes to uS and C_d if the stock price goes to dS. Indeed, the functional form of D and B remains unchanged. To get the new values of D and B, we simply use eq. (1) with the new values of C_u and C_d,

Can we now say, as before, that an opportunity for profitable risk less arbitrage will be available if the current price of the call is not equal to the new value of this portfolio or S – K, whichever is greater? Yes, but there is an important difference. With one period to go, we could plan to lock in a risk less profit by selling an overpriced call and using part of the proceeds to buy the hedging portfolio. At the end of the period, we knew that the market price of the call must be equal to the value of the portfolio, so the entire position could be safely liquidated at that point. But this was true only because the end of the period was the expiration date. Now we have no such guarantee. At the end of the current period, when there is still one period left, the market price of the call could still be in disequilibrium and be greater than the value of the hedging portfolio. If we closed out the position then, selling the portfolio and repurchasing the call, we could suffer a loss which would more than offset our original profit. However, we could always avoid this loss by maintaining the portfolio for one more period. The value of the portfolio at the end of the current period will always be exactly sufficient to purchase the portfolio we would want to hold over the last period. In effect, we would have to readjust the proportions in the hedging portfolio, but we would not have to put up any more money.

Consequently, we conclude that even with two periods to go, there is a strategy we could follow which would guarantee risk less profits with no net investment if the current market price of a call differs from the maximum of ΔS + *Band* S – K. Hence, the larger of these is the current value of the call.

Since Δ and B have the same functional form in each period, the current value of the call in terms of C_u and C_d will again be C = $[pC_u + (1 - p)C_d] / r$ if this is greater than S - K, and C = S - K otherwise. By substituting from eq. (4) into the former expression, and noting that $C_{du} = C_{ud}$, we obtain

$$C=[p^2C_{uu}+2p(1-P)C_{ud}+(1-p)2Cdd] / r^2$$

$$=(p^2 \max[0,u^2S-K] + 2p(1-p)\max[0,duS-K]$$

$$+ (1-p)^2 \max[0, d^2 S - K] / r^2. \qquad (5)$$

A little algebra shows that this is always greater than S - K if, as assumed. r is always greater than one, so this expression gives the exact value of the call. [8]

All of the observations made about formula (3) also apply to formula (5), except that the number of periods remaining until expiration, n, now emerges clearly as an additional determinant of the call value. For formula (5), n = 2. That is, the full list of variables determining C is S, K, n, u, d, and r.

We now have a recursive procedure for finding the value of a call with any number of periods to go. By starting at the expiration date and working backwards, we can write down the general valuation formula for any n:

$$C = \left(\sum_{j=0}^{n} \left\{ \frac{n!}{j!\,(n-j)!} \right\} p^j (1-p)^{n-j} \max[0, u^j d^{n-j} S - K] \right) \Big/ r^n \qquad (6)$$

This gives us the complete formula, but with a little additional effort we can express it in a more convenient way.

Let a stand for the minimum number of upward moves which the stock must make over the next n periods for the call to finish in-the-money. Thus a will be the smallest non-negative integer such that $u^a d^{n-a} S > K$. By taking the natural logarithm of both sides of this inequality, we could write a as the smallest non-negative integer greater than $1og(K/Sd^n) / log(u/d)$.

For all j < a,

$\max[0, u^j d^{n-j} S - K] = 0$,

and for all $j \geq 0$,

$$\max[0, u^j d^{n-j} S - K] = u^j d^{n-j} S - K .$$

Therefore,

$$C = \left(\sum_{j=a}^{n} \left\{ \frac{n!}{j!\,(n-j)!} \right\} p^j (1-p)^{n-j} \max[0, u^j d^{n-j} S - K] \right) \Big/ r^n$$

[8] In the current situation, with no dividends, we can show by a simple direct argument that if there are no arbitrage opportunities, then the call value must always be greater than S – *K* before the expiration date. Suppose that the call is selling for S -*K*. Then there would be an easy arbitrage strategy which would require no initial investment and would always have a positive return. All we would have to do is buy the call, short the stock, and invest *K* dollars in bonds. See Merton (1973). In the general case, with dividends, such an argument is no longer valid. and we must use the procedure of checking every period.

Of course, if $a > n$, the call will finish out-of-the-money even if the stock moves upward every period, so its current value must be zero.

By breaking up C into two terms, we can write

$$C = \left(\sum_{j=a}^{n} \left\{ \frac{n!}{j!\,(n-j)!} \right\} p^j (1-p)^{n-j} \, u^j d^{n-j} \left\{ \frac{u^j d^{n-1}}{r^n} \right\} \right)$$

$$- K \; r^{-n} \left(\sum_{j=a}^{n} \frac{n!}{j!\,(n-j)!} \; p^j (1-p)^{n-j} \right)$$

Now, the latter bracketed expression is the complementary binomial distri-bution function f*[a; n, p]*. The first bracketed expression can also be interpreted as a complementary binomial distribution function *CP[a; n, p']*, where

$$p' \equiv (u / r)p \text{ and } 1 - p' \equiv (d / r)(1 - p).$$

p' is a probability, since $0 < p' < 1$. To see this, note that $p < (r / u)$ and

$$p^j (1-p)^{n-j} \left\{ \frac{(u^j d^{n-j})}{r^n} \right\} = \left(\frac{u}{r} p \right)^j \left(\frac{d}{r} (1-p) \right)^{n-j} = p'^j (1-p')^{n-1}$$

In summary

Binomial Option Pricing Formula

$$C = S\phi[a; n, p'] - Kr^{-n} \; \phi[a; n,p],$$

where

$p ? (r - d) / (u - d)$ and $p' ? (u / r)p$,

a ? the smallest non-negative integer greater than $\log(K/Sd^n)/\log(u/d)$

If $a>n$, $C=0$,

It is now clear that all of the comments we made about the one period evaluation formula are valid for any number of periods. In particular, the value of a call should be the expectation, in a risk-neutral world, of the discounted value of the payoff it will receive. In fact, that is exactly what eq. (6) says, Why, then, should we waste time with the recursive procedure when we can write down the answer in one direct step? The reason is that while: this one-step approach is always technically correct, it is really useful only if we know in advance the circumstances in which a rational individual would prefer to exercise the call before the expiration date. If we do not know this, we have no way to compute the required expectation. In the present example, a call on a stock paying no dividends, it happens that we can determine this information from other sources: the call should never be exercised before the expiration date. As we will see in section 6, with puts or with calls on stocks which pay dividends, we will not be so lucky. Finding the optimal exercise strategy will be an integral part of the valuation problem. The full recursive procedure will then be necessary.

For some readers, an alternative 'complete markets' interpretation of our binomial approach may be instructive. Suppose that p_u and p_d represent the state-contingent discount rates to states *u* and *d*, respectively. Therefore, p_u would be the current price of one dollar received at the end of the period, if and only if state *u* occurs. Each security - a riskless bond, the stock, and the option -must all have returns discounted to the present by p_u and p_d if no risk less arbitrage opportunities are available. Therefore,

$$1 = \pi_u r + \pi_d r,$$

$$S = \pi_u, (uS) + \pi d\,(dS),$$

$$C = \pi_u, C_u + \pi_d C_d$$

The first two equations, for the bond and the stock, imply

$$\pi = \frac{r-d}{u-d} \; \frac{1}{r} \quad \text{and} \quad \pi_d = \frac{u-r}{u-d} \; \frac{1}{r}$$

Substituting these equalities for the state-contingent prices in the last equation for the option yields eq. (3).

It is important to realize that we are not assuming that the riskless bond and the stock and the option are the only three securities in the economy, or that other securities must follow a binomial process. Rather, however these securities are priced in relation to others in equilibrium, among themselves they must conform to the above relationships.

From either the hedging or complete markets approaches, it should be clear that three-state or trinomial stock price movements will not led to an option pricing formula based solely on arbitrage considerations. Suppose, for example, that over each period the stock price could move to *uS* or *dS* or remain the same at S. A choice of D and *B* which would equate the returns in two states could not in the third. That is, a risk less arbitrage position could not be taken. Under the complete markets interpretation, with three equations in now three unknown state-contingent prices, we would lack the redundant equation necessary to price one security in terms of the other two.

Riskless Trading Strategies

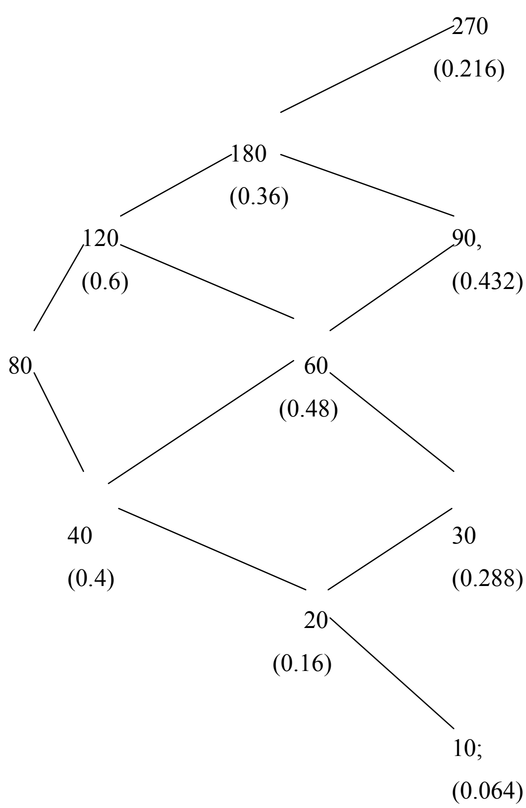
The following numerical example illustrates how we could use the formula if the current market price M ever diverged from its formula value C. If M > C, we would hedge, and if M < C, ‘reverse hedge’, to try and lock in a profit. Suppose the values of the underlying variables are

5=80, n = 3, K =80, u= 1.5, d =0.5, r= 1.1.

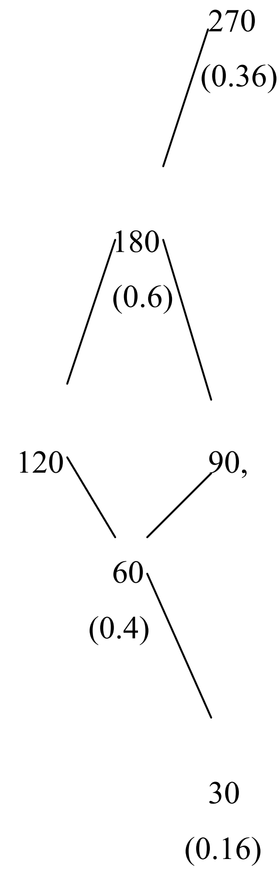
In this case, p= (r – d)/(u – d)=0.6. The relevant values of the discount factor are

r^{-1}=0.909, r^{-2}=0.826, r^{-3}=0.751.

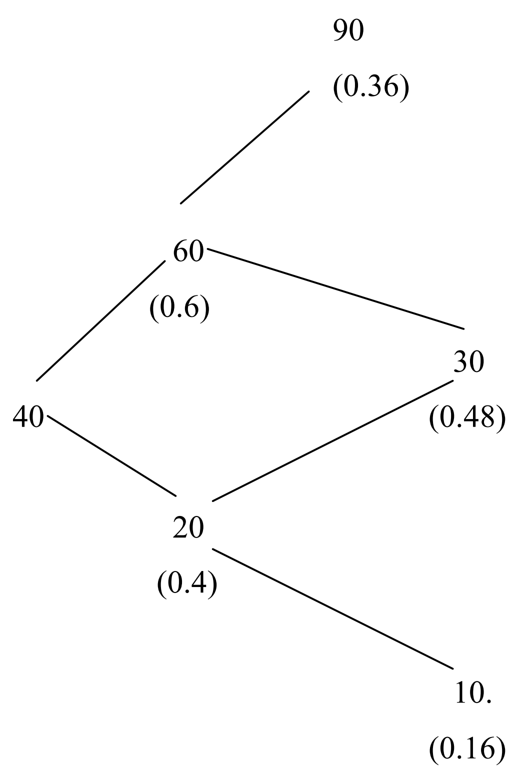
The paths the stock price may follow and their corresponding probabilities (using probability p) are, when n = 3, with S = 80,



when n = 2, if S = 120,



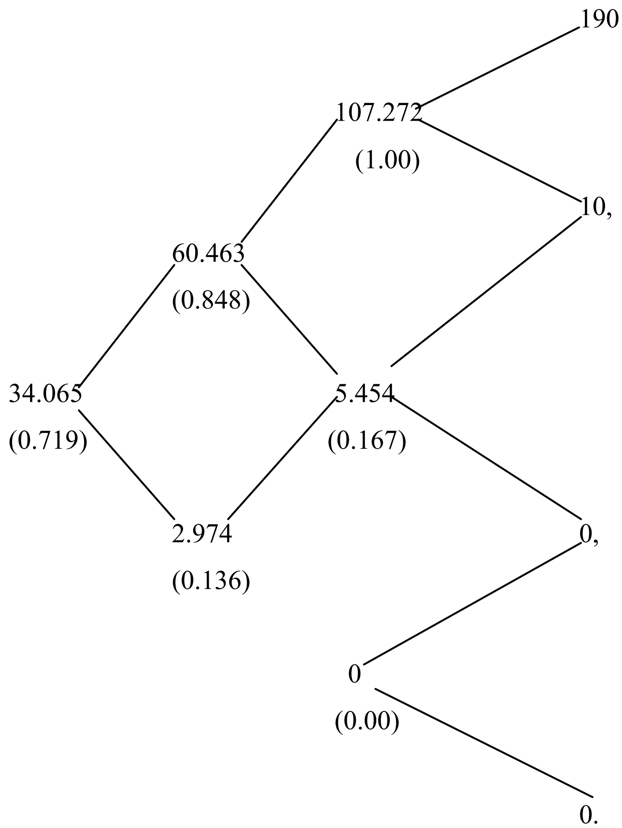
when *n=2*, if S=40,



Using the formula, the current. value of the call would be

C =0.751 [0.064(0) +0.288 (0) + 0.432(90- 80) +0.216(270 - 80)] = 34.065.

Recall that to form a riskless hedge, for each call we sell, we buy and subsequently keep adjusted a portfolio with DSin stock and *B* in bonds, where $\Delta = (C_u - C_d) / (u - d)S$. The following tree diagram gives the paths the call value may follow and the corresponding values of Δ:



With this preliminary analysis, we are prepared to use the formula to take advantage of mispricing in the market. Suppose that when *n* = 3, the market price of the call is 36. Our formula tells us the call should be worth 34.065. The option is over-priced, so we could plan to sell it and assure ourselves of a profit equal to the mispricing differential. Here are the steps you could take for a typical path the stock might follow.

Step 1 (*n* = 3): Sell the call for 36. Take 34.065 of this and invest it in a portfolio containing Δ =0.719 shares of stock by borrowing 0.719(80) - 34.065 = 23.455. Take the remainder, 36 - 34.065 = 1.935, and put it in the bank.*Step* 2 *(n=2):* Suppose the stock goes to 120 so that the new Δ is 0.848. Buy 0.848 - 0.719 = 0.129 more shares of stock at 120 per share for a total expenditure of 15.480. Borrow to pay the bill. With an interest rate of 0.1, you already owe 23.455(1.1) = 25.801. Thus, your total current indebtedness is 25.801 + 15.480 = 41.281.

Step 3 (*n* = 1):" Suppose the stock price low goes to 60. The new t1 is 0.167. Sell 0.848-0.167=0.681 shares at 60 per share, taking in 0.681(60) = 40.860. Use this to pay back part of your borrowing. Since you now owe 41.281(1.1) = 45.409, the repayment will reduce this to 45.409-40.860=4.549.

Step 4d (n=0): Suppose the stock price now goes to 30. The call you sold has expired worthless. You own 0.167 shares of stock selling at 30 per share, for a total value of 0.167 (30) = 5. Sell the stock and repay the 4.549 (1.1) = 5 that you now owe on the borrowing. Go back to the bank and withdraw your original deposit, which has now grown to 1.935 $(1.1)^3$ = 2.575.

Step 4u (n=0): Suppose, instead, the stock price goes to 90. The call you sold is in the money at the expiration date. Buy back the call, or buy one share of stock and let it be exercised, incurring a loss of 90 - 80 = 10 either way. Borrow to cover this, bringing your current indebtedness to 5 +10 = 15. You own 0.167 shares of stock selling at 90 per share, for a total value of 0.167 (90) = 15. Sell the stock and repay the borrowing. Go back to the bank and withdraw your original deposit, which has now grown- to 1.935 $(1.1)^3$ = 2.575.

In summary, if we were correct in our original analysis about stock price movements (which did not involve the unenviable task of predicting whether the stock price would go up or down), and if we faithfully adjust our portfolio as prescribed by the formula, then we can be assured of walking away in the clear at the expiration date, while still keeping the original differential and the interest it has accumulated. It is true that closing out the position before the expiration date, which involves buying back the option at its then current market price, might produce a loss which would more than offset our profit, but this loss could always be avoided by waiting until the expiration date. More-over, if the market price comes into line with the formula value before the expiration date, we can close out the position then with no loss and be rid of the concern of keeping the portfolio adjusted.

It still might seem that we are depending on rational behavior by the person who bought the call we sold. If instead he behaves foolishly and exercises at the wrong time, could he make things worse for us as well as for himself? Fortunately, the answer is no mistakes on his part can only mean greater profits for us. Suppose that he exercises too soon. In that circum-stance, the hedging portfolio will always be worth more than S - *K*, so we could close out the position then with an extra profit.

Suppose, instead, that he fails to exercise when it would be optimal to do so. Again there is no problem. Since exercise is now optimal, our hedging portfolio will be worth S -*K*.[9] If he had exercised, this would be exactly sufficient to meet the obligation and close out the position. Since he did not, the call will be held at least one more period, so we calculate the new values of C_u and C_d and revise our hedging portfolio accordingly. But now the amount required for the portfolio, D*S* + *B*, is less than the amount we have available, S -*K*. We can withdraw these extra profits now and still maintain the hedging portfolio. The longer the holder of the calI goes on making mistakes, the better off we will be.

[9]If we were reverse hedging by buying an undervalued call and selling the hedging portfolio, then we would ourselves want to exercise at this point. Since we will receive S - *K* from exercising, this will be exactly enough money to buy back the hedging portfolio.

Consequently, we can be confident that things will eventually work out right no matter what the other party does. The return on our total position, when evaluated at prevailing market prices at intermediate times, may be negative. But over a period ending no later than the expiration date, it win be positive.

In conducting the hedging operation, the essential thing was to maintain the proper proportional relationship: for each call we are short, we hold J shares of stock and the dollar amount B in bonds in the hedging portfolio. To emphasize this, we will refer to the number of shares held for each call as the hedge ratio. In our example, we kept the number of calls constant and made adjustments by buying or selling stock and bonds. As a result, our profit was independent of the market price of the call between the time we initiated the hedge and the expiration date. If things got worse before they got better, it did not matter to us.

Instead, we could have made the adjustments by keeping the number of shares of stock constant and buying or selling calls and bonds: However, this could be dangerous. Suppose that after initiating the position, we needed to increase the hedge ratio to maintain the proper proportions. This can be achieved in two ways:

a. Buy more stock, or

b. Buy back some of the calls.

If we adjust through the stock, there is no problem. If we insist on adjusting through the calls, not only is the hedge no longer risk less, but it could even end up losing money! This can happen if the call has become even more overpriced. We would then be closing out part of our position in calls at a loss. To remain hedged, the number of calls we would need to buy back depends on their value, not their price. Therefore, since we are uncertain about their price, we then become uncertain about the return from the hedge.

Worse yes, if the call price gets high enough, the loss on the closed portion of our position could throw the hedge operation into an overall loss.

To see how this could happen, let us rerun the hedging operation, where we adjust the hedge ratio by buying and selling calls.

Step 1 (n = 3): Same as before.

Step 2 (n =2): Suppose the stock goes to 120, so that the new J =0.848. The call price has gotten further out of line and is now selling for 75. Since its value is 60.463, it is now overpriced by 14.537. With 0.719 shares you must buy back 1-0.848=0.152 calls to produce a hedge ratio of 0.848 =0.719/0.848. This costs 75(0.152) =11.40. Borrow to pay the bill. With the interest rate of 0.1, you already owe 23.455(1.1) = 25.801. Thus, your total current indebtedness is 25.801 + 11.40 = 37.201.

Step 3 (n = 1): Suppose the stock goes to 60 and the call is selling for 5.454. Since the call is now fairly valued, no further excess profits can be made by continuing to hold the position. Therefore, liquidate by selling your 0.719 shares for 0.719(60) = 43.14 and close out the call position by buying back 0.848 calls for 0.848(5.454) = 4.625. This nets 43.14-4.625=38.515. Use this to pay back part of your borrowing. Since you now owe 37.200.1) = 40.921, after repayment you owe 2.406. Go back to the bank and withdraw your original deposit, which has now grown to $1.935(1.1)^2$ = 2.341. Unfortunately, after using this to repay your remaining borrowing, you still owe 0.065.

Since we adjusted our position at Step 2 by buying overpriced calls, our profit is reduced. Indeed, since the calls were considerably overpriced, we actually lost money despite apparent profitability of the position at Step 1. We can draw the following adjustment rule from our experiment: To adjust a hedged position, never buy an overpriced option or sell an under priced option. As a corollary, whenever we can adjust a hedged position by buying more of an underpriced option or selling more of an overpriced option, our profit will be enhanced if we do so. For example, at Step 3 in the original hedging illustration, had the call still been overpriced, it would have been better to adjust the position by selling more calls rather than selling stock. In summary, by choosing the right side of the position to adjust at intermediate dates, at a minimum we can be assured of earning the original differential and its accumulated interest, and we may earn considerably more.

Limiting Cases

In reading the previous sections, there is a natural tendency to associate with each period some particular length of calendar time, perhaps a day. With this in mind, you may have had two objections. In the first place, prices a day from now may take on many more than just two possible values. Furthermore, the market is not open for trading only once a day, but, instead, trading takes place almost continuously.

These objections are certainly valid. Fortunately, our option pricing approach has the flexibility to meet them. Although it might have been natural to think of a period as one day, there was nothing that forced us to do so. We could have taken it to be a much shorter interval – say an hour -or even a minute. By doing so, we have met both objections simultaneously. Trading would take place far more frequently, and the stock price could take on hundreds of values by the end of the day.

However, if we do this, we have to make some other adjustments to keep the probability small that the stock price will change by a large amount over a minute. We do not want the stock to have the same percentage up and down moves for one minute as it did before for one day. But again there is no need for us to have to use the same values. .We could, for example, think of the price as making only a very small percentage change over each minute.

To make this more precise, suppose that h represents the elapsed time between successive stock price changes. That is, if t is the fixed length of calendar time to expiration, and n is the number of periods of length h prior to expiration, then

$$h=t/n$$

As trading takes place more and more frequently, h gets closer and closer to zero. We .must then adjust the interval-dependent variables r, u, and d in such a way that. we obtain empirically realistic results as h becomes smaller, or, equivalently, as $\to \infty$

When we were thinking of the periods as having a fixed length, r represented both the interest rate over a fixed length of calendar time and the interest rate over one period. Now we need to make a distinction between these two meanings. We will let r continue to mean one plus the interest rate over a fixed length of calendar time. When we have occasion to refer to one plus the interest rate over a period (trading interval) of length *h*, we will use the symbol .

Clearly, the size of depends on the number of subintervals, *n,* into which *t* is divided. Over the n periods until expiration, the total return is ^{n}, where n = *t/h.* Now not only do we want to depend on *n,* but we want it to depend on 11 in a particular way - so that as *n* changes the total return over the fixed time *t.* remains the same. This is because the interest rate obtainable over some fixed length of calendar time should have nothing to do with how we choose to think of the length of the time interval *h.*

If r (without the 'hat') denotes one plus the rate of interest over a *fixed* unit, of calendar time, then over elapsed time *t,* r' is the total return.[10] Observe that this measure of total return does not depend on *n.* As we have argued, we want to choose the dependence of on *n,* so that

$$^{n} = r^{t},$$

for any choice of *n.* Therefore, $= r^{t/n}$. This last equation shows how must depend on *n* for the total return over elapsed time *t* to be independent of *n.*

We also need to define *u* and *d* in terms of *n.* At this point, there are two significantly different paths we can take. Depending on the definitions we choose, as $n \rightarrow \infty$ *('%* (or, equivalently, as $h \rightarrow 0$ *),* we can have either a continuous or a jump stochastic process. In the first situation very small random changes in the stock price will be occurring in each very small time interva1. The stock price will fluctuate incessantly, but its path can be drawn without lifting pen from paper. In contrast, in the second case, the stock price will usually move in a smooth deterministic way, but will occasionally experience sudden discontinuous changes. Both can be derived from our binomial process simply by choosing how *u* and *d* depend on *n.* We examine in detail only the continuous process which leads 'to the option pricing formula originally derived by Fischer Black and Myron Scholes, Subsequently, we indicate how to develop the jump process formula originally derived by John Cox and Stephen Ross.

[10]The. scale of this unit (perhaps a day, or a year) is unimportant as long as r and *I* are expressed in the same scale.

Recall that we supposed that over each period the stock price would experience a one plus rate of return of *u* with probability *q* and *d* with probability 1 - *q.* It will be easier and clearer to work, instead, with the natural logarithm of the one plus rate of return, log *u* or log *d.* This gives the continuously compounded rate of return on the stock over each period. It is a random variable which, in each period, will be equal to log *u* with probability *q* and log *d* with probability 1 - *q.*

Consider a typical sequence of five moves, say *u, d, u, u, d.* Then the final stock price will be $S^{*}=uduudS$; $S^{*}/S=u^{3}d^{2}$, and $\log(S^{*}/S)=3\log u + 2\log d$, More generally, over *n* periods,

$$\text{Log}\,(S^{*}/S)= j\log u + (n-j)\log d = j\log(u/d)+n\log d,$$

where j is the (random) number of upward moves occurring during the 11 periods to expiration. Therefore, the expected value of log (S* / S) is

$$E[\log(S^{*}/S)] = \log(u/d)\;E(j) + n\log d,$$

and its variance is

$$\text{var}\,[\log(S^{*}/S)] = [\log(u/d)]^{2}\,.\,\text{var}(j).$$

Each of the *n* possible upward moves has probability *q.* Thus, $E(j)=nq$. Also, since the variance each period is $q(1-q)^{2} + (1-q)(O-q)^{2} = q(1-q)$, then var (j) = nq (1 – q). Combining all of this, we have

$$E\,[\log(S^{*}/S)] = [q\log(u/d) + \log d] \equiv mn$$

$$\text{var}[\log(S^{*}/S)] = q(1-q)[\log(u/d)]^{2}n \equiv sn$$

Let us go back to our discussion. We were considering dividing up our original longer time period (a day) into many shorter periods (a minute or even less). Our procedure calls for, over fixed length of calendar time *t* making *n* larger and larger. Now if we held everything else constant while we let *n* become large, we would be faced with the problem we talked about earlier. In fact, we would certainly not reach a reasonable conclusion if either m *n* or s went to zero or infinity as *n* became large. Since *t* is a fixed length of time, in searching for a realistic result, we must make the appropriate adjustments in *u. d,* and *q.* In doing that, we would at least want the mean and variance of the continuously compounded rate of return of the assumed stock price movement to coincide with that of the actual stock price as $\rightarrow \infty$ Suppose we label the actual empirical values of m*n* and $s^{2}n$ as μt and $s^{2}t$, respectively. Then we would want to choose *u, d,* and *q,* so that

$$[q\log(u/d)+\log d] \rightarrow \mu t$$

$$\text{as } n \rightarrow \infty$$

$$q(1-q)[\log(u/d)]^{2} \rightarrow \sigma$$

A little algebra shows we can accomplish this by letting

$$u = e^{s\sqrt{t/n}}\;,\quad d = e^{-\sigma\sqrt{t/n}}\;,\quad q = \tfrac{1}{2} + \tfrac{1}{2}(\mu/s)\sqrt{t/n}$$

In this case, for any n

$$\mu n = \mu t \text{ and } \sigma^{2}n = [\sigma^{2} - \mu^{2}(t/n)]t$$

Clearly, as $n \rightarrow$ ¥, $s^{2}n$ *à* $s^{2}n$, while $\mu n = \mu t$ for all values of *n.*

Alternatively, we could have chosen *u, d,* and *q* so that the mean and variance of the future stock price for the discrete binomial process approach the pre specified mean and variance of the actual stock price as $n \rightarrow \infty$. However, just as we would expect, the same values will accomplish this as well. Since this would not change our conclusions, and it is computationally more convenient to work with the continuously compounded rates of return, we will proceed in that way.

This satisfies our initial requirement that the limiting means and variances coincide, but we still need to verify that we are arriving at a sensible limiting probability distribution of the continuously compounded rate of return. The mean and variance only describe certain aspects of that distribution. For our model, the random" continuously compounded rate of return over a period of length *t* is the sum of *n* independent random variables, each of which can take the value log *u* with probability *q* and log *d* with probability 1 - *q.* We wish to know about the distribution of this sum as *n* becomes large and *q, u,* and *d* are chosen in the way described. We need to remember that as we change *n,* we are not simply adding one more random variable to the previous sum, but instead are changing the probabilities and possible outcomes for every member of the sum. At this point, we can rely on a form of the central limit theorem which, when applied to our problem, says that, as $n \to ¥$, if

$$\frac{[\log u - \hat{\mu}3] + (1- q)\,[\log d - \hat{\mu}]^3}{\hat{\sigma}^3 \sqrt{n}} \to 0$$

then

$$\left(\text{Prob}\left\{ \frac{\log(S^*/S) - \hat{\mu} n}{\sigma\sqrt{n}} \right\} \bullet z \right) \to N(z)$$

where *N (z)* is the standard normal distribution functio'1. Putting this into words, as the number of periods into which the fixed length of time to expiration is divided approaches infinity, the probability that the standar-dized continuously compounded rate of return of the stock through the expiration date is not greater than the number z approaches the probability under a standard normal distribution.

The initial condition says roughly that higher-order properties of the distribution, such as how it is skewed, become less and less important, relative to its standard deviation, as $n \to ¥$. We can verify that the condition is satisfied by making the appropri-ate substitutions and finding

$$\frac{q\,[\log u - \hat{\mu}]^3 + (1-q)\,[\log d - \hat{m}]^3 - (1-q)^2 + q^2}{\hat{s}^3 Ö n}$$

then

$$\text{Prob}\left(\left\{ \frac{\log(S^*/S) - \mu n}{\hat{\sigma}\sqrt{n}} \right\} ?\, z \to N(z), \right)$$

where N (z) is the standard normal distribution function putting this into words as the number of periods into which the fixed length of time to expiration is divided approaches infinity, the probability that the standardized continuously compounded rate of return of the stock through the expiration date is not greater than the number z approaches the probability under a standard normal distribution.

The initial condition says roughly that higher-order properties of the distribution, such as how it is skewed, become less and less important, relative to its standard deviation, as $n \to \infty$. We can verify that the condition is satisfied by making the appropri-ate substitutions and finding

$$\frac{q\,[\log u - \mu]^3 + (1-q)\quad[\log d - \mu]^3}{\hat{\sigma}^3\sqrt{n}} = \frac{(1-q)^2 + q^2}{\sqrt{nq\,(1-q)}}$$

which goes to zero as $n \to \infty$ since $q = ½ + ½\,(\mu/s)\,Ö\,t/n$ Thus, the multiplicative binomial model for stock prices includes the lognormal distribution as a limiting case.

Black and' Scholes began directly with continuous trading and the assumption of a lognormal distribution for stock prices. Their approach relied on some quite advanced mathematics. However, since our approach contains continuous trading and the lognormal distribution as a limiting case, the two resulting formulas should then coincide. We will see shortly that this is indeed true, and we will have the advantage of using a much simpler method. It is important to remember, however, that the economic arguments we used to link the option value and the stock price are exactly the same as those advanced by Black and Scholes (1973) and Merton (1973, 1977.

The formula derived by Black and Scholes, rewritten in terms of our notation, is

Black -Scholes Option Pricing Formula

$$C = SN(x) - Kr^{-1} N(x - sÖ\overline{t}),$$

where

$$\frac{\log(S/Kr^{-t})}{sÖ t} + ½\, sÖ\overline{t}$$

We now wish to confirm that our binomial formula converges to the Black-Scholes formula when *t* is divided into more and more subintervals, and ,*u*, *d.* and *q* are chosen in the way we described that is, in a way such that the multiplicative binomial probability distribution of stock prices goes to the lognormal distribution.

For easy reference, let us recall our binomial option pricing formula:

$C = Sf\,[a; n, p'] - K^{-n}\; f\,[a; n, p].$

The similarities are readily apparent. $^{-n}$ is, of course, always equal to r^{1}. Therefore, to show the two formulas conve_ge, we need only show that as $n \to \infty$.

$f[a; n, p'] \to N(x)$ and $f[a; n, p] \to N(x - sÖ\, t).$

We will consider only $f\,[a;n,p]$, since the argument is exactly the same for $f[a; n, p']$.

The complementary binomial distribution function $f\,[a;n,p]$ is the prob-ability that the sum of *n* random variables, each of which can take on the value I with probability *p* and 0 with probability 1 - *p*, will be greater than or equal to *a*. We know that the random value of this sum, j, has mean *np* and standard deviation $Önp\,(1 - p)$. Therefore,

$$1 - f\,[a; n, p] = \text{Prob}[\,j \bullet a - 1]$$

$$= \text{Prob}\, \frac{j - np}{Önp(1-p)} \;?\; \frac{a - 1 - np}{Önp(1-p)}$$

Now we can make an analogy with our earlier discussion. If we consider a stock which in each period will move to *uS* with probability *p* and *dS* with probability 1 – p, then $\log(S^*/S) = j\,\log(u/d) + n\,\log d$. The mean and variance of the continu-ously compounded rate of return of this stock are

$m_p = p\,\log(u/d) + \log d$ and $s^2_p = p(1-p)[\log(u/d)]^2$.

Using these equalities, we find that

$$\frac{j - np}{np\,(1-p)} \qquad \frac{\log (S^*/S) - m_p n}{\hat{s}_p Ö\overline{n}}$$

Recall from the binomial formula that

a – I = log (K / Sd^n) / log(u / d) - e

= [log (K/S) – n log d] log (u/d) - e

where e is a number between zero and one. Using this and the definitions of m_p and s^2_p, with a little algebra, we have

$$\frac{a - 1 - np}{np\,(1-p)} = \frac{\log (K/S) - \hat{\mu}_p n - e \log (u/d)}{\hat{s}_p Ö\overline{n}}$$

Putting these results together,

1 - f [a; n, p]

$$= \text{Prob}\left(\frac{\log (S^*/S) - \mu_p n}{\hat{\sigma}_p \sqrt{n}} \bullet \frac{\log (K/S) - \mu_p n - \varepsilon \log (u\ d)}{\hat{\sigma}_p \sqrt{n}}\right)$$

We are now in a position to apply the central limit theorem. First, we must check if the initial condition,

$$\frac{p[\log u - \mu_p]^3 + (1-p)[\log d - \mu_p]^3}{\hat{s}_p Ö\overline{n}} = \frac{(1-p)^{2+} + p^2}{Ö\overline{np}\,(1-p)} \to 0$$

as $n \to ¥$, is satisfied. By first recalling that $p \equiv (\bullet - d) / (u-d)$, and then $= r^{t/n}, u = e^{sÖ t/n}$ and $d = e^{sÖ t/n}$, it is possible to show that as $n \to ¥$,

$P \to ½ + ½ (\log r – ½\sigma^2 / s) \sqrt{t/n}$

As a result, the initial condition holds, and we are justified in applying the central limit theorem.

To do so, we need only evaluate $\hat{\mu}_p n$, $\hat{\sigma}^2_p n$ and $\log(u/d)$ as $n \to \infty$.[11] Examination of our discussion for parameterizing q shows that as $n \to \infty$,

$$\hat{\mu}_p n \to (\log r - \tfrac{1}{2}\sigma^2)t \quad \text{and} \quad \hat{\sigma}_p \sqrt{n} \to \sigma\sqrt{t}.$$

Furthermore, $\log(u/d) \to 0$ as $n \to \infty$.

For this application of the central limit theorem, then, since

$$\frac{\log(K/S) - \hat{\mu}_p n - \varepsilon \log(u/d)}{\hat{\sigma}_p \sqrt{n}} \to z = \frac{\log(K/S) - (\log r - \tfrac{1}{2}\sigma^2)t}{\sigma\sqrt{t}},$$

we have

$$1 - \Phi[a; n, p] \to N(z) = N\left[\frac{\log(Kr^{-t}/S)}{\sigma\sqrt{t}} + \tfrac{1}{2}\sigma\sqrt{t}\right].$$

The final step in the argument is to use the symmetry property of the standard normal distribution that $1 - N(z) = N(-z)$. Therefore, as $n \to \infty$,

$$\Phi[a; n, p] \to N(-z) = N\left[\frac{\log(S/Kr^{-t})}{\sigma\sqrt{t}} - \tfrac{1}{2}\sigma\sqrt{t}\right] = N(x - \sigma\sqrt{t}).$$

Since a similar argument holds for $\Phi[a; n, p']$, this completes our dem-

[11] A surprising feature of this evaluation is that although $p \neq q$ and thus $\hat{\mu}_p \neq \hat{\mu}$, and $\hat{\sigma}_p \neq \hat{\sigma}$, nonetheless $\hat{\sigma}_p \sqrt{n}$ and $\hat{\sigma}\sqrt{n}$ have the same limiting value as $n \to \infty$. By contrast, since $\mu \neq \log r - (\tfrac{1}{2}\sigma^2)$, $\hat{\mu}_p n$ and $\hat{\mu} n$ do not. This results from the way we needed to specify u and d to obtain convergence to a lognormal distribution. Rewriting this as $\sigma\sqrt{t} = (\log u)\sqrt{n}$, it is clear that the limiting value σ of the standard deviation does not depend on p or q, and hence must be the same for either. However, at any point before the limit, since

$$\hat{\sigma}^2 n = \left(\sigma^2 - \mu^2 \frac{t}{n}\right)t \quad \text{and} \quad \hat{\sigma}^2_p n = \left[\sigma^2 - (\log r - \tfrac{1}{2}\sigma^2)^2 \frac{t}{n}\right]t,$$

$\hat{\sigma}$ and $\hat{\sigma}_p$ will generally have different values.

The fact that $\hat{\mu}_p n \to (\log r - \tfrac{1}{2}\sigma^2)t$ can also be derived from the property of the lognormal distribution that

$$\log E[S^*/S] = \mu_p t + \tfrac{1}{2}\sigma^2 t,$$

where E and μ_p are measured with respect to probability p. Since $p \equiv (\hat{r} - d)/(u - d)$, it follows that $\hat{r} = pu + (1-p)d$. For independently distributed random variables, the expectation of a product equals the product of their expectations. Therefore,

$$E[S^*/S] = [pu + (1-p)d]^n = \hat{r}^n = r^t.$$

Substituting r^t for $E[S^*/S]$ in the previous equation, we have

$$\mu_p = \log r - \tfrac{1}{2}\sigma^2.$$

onstration that the binomial option pricing formula contains the Black–Scholes formula as a limiting case.[12,13]

As we have remarked, the seeds of both the Black–Scholes formula and a continuous-time jump process formula are both contained within the binomial formulation. At which end point we arrive depends on how we take limits. Suppose, in place of our former correspondence for u, d, and q, we instead set

$$u=u, \qquad d=e^{\zeta(t/n)}, \qquad q=\lambda(t/n).$$

This correspondence captures the essence of a pure jump process in which each successive stock price is almost always close to the previous price ($S \to dS$), but occasionally, with low but continuing probability, significantly different ($S \to uS$). Observe that, as $n \to \infty$, the probability of a change by d becomes larger and larger, while the probability of a change by u approaches zero.

With these specifications, the initial condition of the central limit theorem we used is no longer satisfied, and it can be shown the stock price movements converge to a log-Poisson rather than a lognormal distribution as $n \to \infty$. Let us define

$$\Psi[x; y] \equiv \sum_{i=x}^{\infty} \frac{e^{-y} y^i}{i!},$$

[12]The only difference is that, as $n \to \infty$, $p' \to \frac{1}{2} + \frac{1}{2}[(\log r + \frac{1}{2}\sigma^2)/\sigma]\sqrt{t/n}$.

Further, it can be shown that as $n \to \infty$, $\Delta \to N(x)$. Therefore, for the Black–Scholes model, $\Delta S = SN(x)$ and $B = -Kr^{-t}N(x - \sigma\sqrt{t})$.

[13]In our original development, we obtained the following equation (somewhat rewritten) relating the call prices in successive periods:

$$\left(\frac{\hat{r}-d}{u-d}\right)C_u + \left(\frac{u-\hat{r}}{u-d}\right)C_d - \hat{r}C = 0.$$

By their more difficult methods, Black and Scholes obtained directly a partial differential equation analogous to our discrete-time difference equation. Their equation is

$$\tfrac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + (\log r) S \frac{\partial C}{\partial S} - \frac{\partial C}{\partial t} - (\log r) C = 0.$$

The value of the call, C, was then derived by solving this equation subject to the boundary condition $C^* = \max[0, S^* - K]$.

Based on our previous analysis, we would now suspect that, as $n \to \infty$, our difference equation would approach the Black–Scholes partial differential equation. This can be confirmed by substituting our definitions of $\hat{r}$, u, d in terms of n in the way described earlier, expanding C_u, C_d in a Taylor series around $(e^{\sigma\sqrt{h}}S, t-h)$ and $(e^{-\sigma\sqrt{h}}S, t-h)$, respectively, and then expanding $e^{\sigma\sqrt{h}}$, $e^{-\sigma\sqrt{h}}$, and r^h in a Taylor series, substituting these in the equation and collecting terms. If we then divide by h and let $h \to 0$, all terms of higher order than h go to zero. This yields the Black–Scholes equation.

as the complementary Poisson distribution function. The limiting option pricing formula for the above specifications of *u*, *d*, and *q* is then

Jump Process Option Pricing Formula

$C = ?\ [x: y] - Kr^{-1}\ ?\ [x; y / u]$,
where

$y\ ?\ (\log r -\ ?)\ ut/(u - 1)$,

and
$x\ ?$ the smallest non-negative integer greater than $(\log(K/S) -\ ?\ t / \log u$.

A very similar formula holds if we let $u = e^{?(t/n)}$, $d = d$, and $1 - q = ?(t/n)$.

Dividends and Put Pricing

So far we have been assuming that the stock pays no dividends. It is easy to do away with this restriction. We will illustrate this with a specific dividend policy: the stock maintains a constant yield, δ, on each ex-dividend date. Suppose there is one period remaining before expiration and the current stock price is S. If the end of the period is an ex-dividend date, then an individual who owned the stock during the period will receive at that time a dividend of either δuS or δdS. Hence, the stock price at the end of the period will be either $u(1 - \delta)$ S, or $d(1 - \delta)$ S, where v = 1 if the end of the period is an ex-dividend date and v = 0 otherwise. Both δ and v are assumed to be known with certainty.

When the call expires, its contract and a rational exercise policy imply that its value must be either

$C_u = \max\ [0, u(1\text{-d})^v S - K]$,

or

$C_d = \max\ [0, d(1 - \text{d})^v S - K]$.

Therefore,

$$C \begin{cases} C_u = \max\ [0, u(1 - \text{d})^v S - K], \\ C_d = \max\ [0, d(1 - \text{d})^v S - K] \end{cases}$$

Now we can proceed exactly as before. Again we can select a portfolio of Δ shares of stock and the dollar amount B in bonds which will have the same end-of-period value as the call.[14] By retracing our previous steps, we can show that

$C = [pC_u + (1 + P)\ C_d /\ \ ,$

If this is greater than S – K and C=S – K otherwise. Here, once again = (-d) / (u – d) and D = ($C_u - C_d$) / (u – d) / S.

Thus far the only change is that $(1 - \delta)^v$ S in the values for C_u and C-$_d$. Now we come to the major difference: early exercise may be optimal. To see this, suppose that v = 1 and d (1- δ) S > K. Since u > d, then, also, u (1 - δ) S > K. In this case, C_u = u (1 - δ) S - K and C_d = d (1 - δ) S - K. Therefore, since (u /) p + (d /) (1 – p) = 1, [pC_u + (1 – p) C_d] / = (1 - δ) S - (K /). For sufficiently high stock prices, this can obviously b less than S – K. Hence, there are definitely some circumstances in which no one would be willing to hold the call for one more period.

In fact, there will always be a critical stock price, S such that if S >, the call should be exercised immediately. \ will be the stock price at which [pC_{u+} (1 – p)C_d] / = S – K. [15] That is it is the lowest stock price at which the value of the hedging portfolio exactly equals S – K. This means \ will, other things equal, be lower the striking price.

We can extend the analysis to an arbitrary number of periods in the same way as before. There is only one additional difference, a minor modification in the heading operation. Now the funds in the hedging portfolio will be increased by any dividends received, or decreased by the restitution required for dividends paid while the stock is held short.

Although the possibility of optimal exercise before the expiration date causes no conceptual difficulties, it does seem to prohibit a simple closed form solution for the value of a call with many periods to go. However, our analysis suggests a sequential numerical procedure which will allow us to calculate the continuous time value to any desired degree of accuracy.

Let C be the current value of a call with n periods remaining. Define

$$\underline{v}(n, i) = \sum_{k=1}^{n-i} vk'/$$

So that $\overline{v}\,(n - i)$ is the number of ex-dividend dates occurring during the next $n - i$ periods from now, given that the current stock price S has changed to $u^j d^{\,n-i-j}(1 - d)^{v(n-i)}S$, where
$j = 0,1,2,...,n - i$.

With this notation, we are prepared to solve for the current value of the call by working backward in time from the expiration date. At expiration, $i = 0$, so that

$C(n, 1, j) = \max[0, u^j d^{n-j} - (1 - d)^{v(n,0)} S - K]$ for j =0, 1,..., n.
One period before the expiration date, $i = 1$ so that

$$C(n, 1, j) = \max[u^J d^{n-1-j}(1 - d)^{v(n,1)} S - K,$$
$$[pC(n, 0, j + 1) + (1 - p)\, C(n, 0, j)] / ? \;]$$
$$\text{for } j=0, 1,...,n - 1 .$$

More generally, i periods before expiration

$$C(n, i, j) = \max[u^j d^{n-1-j}(1 - d)^{v(n,i)} S - K,$$
$$[pC(n, i-1, j + 1) + (1 - p)C(n, i-1, j)] / ? \;]$$
$$\text{for } j = 0,1,\ldots, n - i$$

Observe that each prior step provides the inputs needed to evaluate the right-hand arguments of each succeeding step. The number of calculations decreases as we move backward in time. Finally, with n periods before expiration, since $i = n$,

$$C = C(n, n, 0) = \max[S - K, [pC(n, n-1, 1) + (1 - p)C(n, n-1, 0)] / ? \;],$$

and the hedge ratio is

$$\Delta = \frac{C(n, n-1, 1) - C(n, n-1, 0)}{(u - d)S}.$$

We could easily expand the analysis to include dividend policies in which the amount paid on any ex-dividend date depends on the stock price at that time in a more general way.[16] However; this will cause some minor complications. In our present example with a constant dividend yield, the possible stock prices n - *i* periods from now are completely determined by the total number of upward moves (and ex-dividend dates) occurring during that interval. With other types of dividend policies, the enumeration will be more complicated, since then the terminal stock price will be affected by the timing of the upward moves as well as their total number. But the basic principle remains the same. We go to the expiration date and calculate the call value for all of the possible prices that the stock could have then. Using this information, we step back one period and calculate the call values for all possible stock prices at that time, and so forth.

We will now illustrate the use of the binomial numerical procedure in approximating continuous-time call values. In order to have an exact continuous-time formula to use for comparison, we will consider the case with no dividends. Suppose that we are given the inputs required for the Black-Scholes option pricing formula: S, *K, t, s,* and *r.* To convert this information into the inputs *d, u,* and required for the binomial numerical procedure, we use the relationships:

$d = 1 / u, \qquad u = e^{s\, Öt/n} \qquad = r^{t/n}$

Table 2 gives us a feeling for how rapidly option values approximated by the binomial method approach the corresponding limiting Black-Scholes values given by $n = \infty$. At $n = 5$, the values differ by at most \$0.25; and at $n = 20$, they differ by at most \$0.07. Although not shown, at $n = 50$, the greatest difference is less than \$0.03, and at $n = 150$, the values are identical to the penny.

To derive a method for valuing puts, we again use the binomial for-mulation. Although it has been convenient to express the argument in terms of a particular security, a call, this is not essential in any way. The same basic analysis can be applied to puts.

Letting P denote the current price of a put, with one period remaining before expiration, we have

$$P \;\begin{cases} P_u = \max\,[0, K - u\,(1 - d)^v S\,], \\ P_d = \max\,[O, K - d\,(1 - d)^v S]. \end{cases}$$

Once again, we can choose a portfolio with DS in stock and *B* in bonds which will have the same end-of-period values as the put. By a series of steps which are formally equivalent to the ones, we followed in section 3, we can show that

$P = [pP_u + (1 - p)\,P_d\,] / \quad ,$

if this is greater than *K* - S, and P =*K* - S otherwise. As before, $p = (\; - d) / (Pu\text{-}Pd)$ and $A = (P_u\text{-}P_d) / (u - d)\, S$ . . Note that for puts, since $P_u \quad d_d$, then $\Delta \quad 0$ This means that if we sell an overvalued put, the hedging portfolio which we buy will involve a short position in the stock.

We might hope that with puts we will be spared the complications caused by optimal exercise before the expiration date. Unfortunately, this is not the case. In fact, the situation is even worse in this regard. Now there are always some possible circumstances in which no one would be willing to hold the put for one more period.

To see this, suppose $K > u\,(1 - \delta)^v S$. Since $u > d$, then, also, $K > d(1 - \delta)^v S$. In this case, $P_u = K\text{-}u\,(1 - \delta)^v S$ and $P_d = K - d\,(1 - \delta)^v S$. Therefore, since $(u/ \;)p + (d/ \;)(1 - p) = 1$,

$[pP_u + (1 - p)P_d] / \quad = (K/ \;) - (1\text{-}\,\delta)^v\, S.$

If there are no dividends (that is, $v = 0$), then this is certainly less than *K* - S. Even with $v = 1$, it will be less for a sufficiently low stock price.

Thus, there will now be a critical stock price, \ , such that if S < \ , the put should be exercised immediately. By analogy with our discussion for the call, we can see that this is the stock price at which $[pP_u + (1 - p)P_d] / \quad = K - S$. Other things equal, S will be higher the lower the dividend yield, the higher the interest rate, and the higher the striking price. Optimal early exercise thus becomes more likely if the put is deep-in-the-money and the interest rate is high. The effect of dividends yet to be paid diminishes the advantages of immediate exercise, since the put buyer will be reluctant to sacrifice the forced declines in the stock price on future ex-dividend dates.

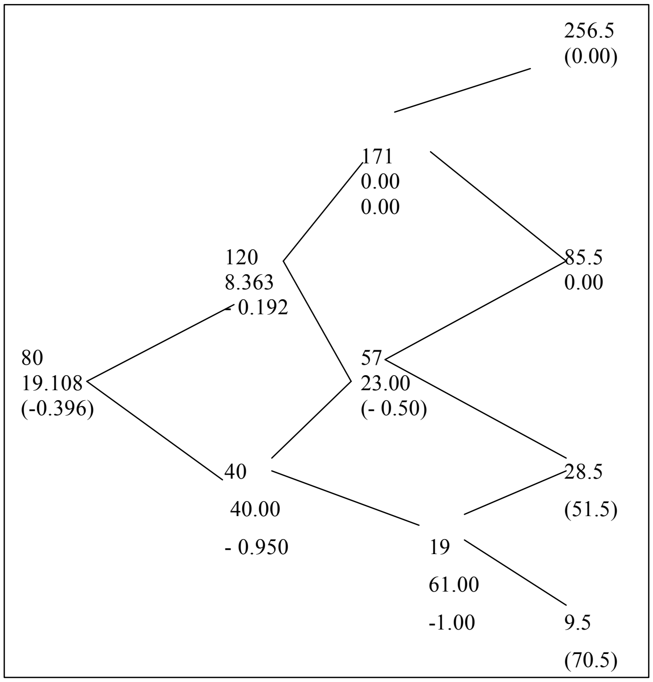
This argument can be extended in the same way as. before to value puts with any number of periods to go. However, the chance for optimal exercise before the expiration date once again seems to preclude the possibility of expressing this value in a simple form. But our analysis also indicates that, with slight modification, we can value puts with the same numerical techniques' we use for calls. Reversing the difference between the stock price and the striking price at each stage is the only change.[17]

[17] Michael Parkinson (1977) has suggested a similar numerical procedure based on a trinomial process, where the stock price can either increase, decrease, or remain unchanged. In fact, given the theoretical basis for the binomial numerical procedure provided, the numerical method can be generalized to permit $k + 1 \bullet n$ jumps to new stock prices in each period. We can consider exercise only every *k* periods, using the binomial formula to leap across intermediate periods. In effect, this means permitting $k + 1$ possible new stock prices before exercise is again considered. That is, instead of considering exercise *n* times, we would only consider it about n/k times. For fixed *t* and *k,* as $n \to ¥$, option values will approach their continuous-time values.
This alternative procedure is interesting, since it may enhance computer efficiency. At one extreme, for calls on stocks which do not pay dividends setting $k + I = n$ gives the most efficient results. However, when the effect of potential early exercise is important and greater accuracy is required. the most efficient results are achieved by setting $k = 1$ as in our description above.

The diagram presented in table 3 shows the stock prices, put values, and values of D obtained in this way for the example given in section 4. The values used there were *S*=80, *K=80, n=3, u=1.5, d=0.5,* and $= 1.1$. To include dividends as well, we assume that a cash dividend of five percent ($\delta = 0.05$) will be paid at the end of the last period before the expiration date.

Thus, $(1 - \delta)^{v(n,0)} = 0.95$, $(1- \delta)^{v(n,1)} = 0.95$, and $(1 -\delta)^{v(n,2)} = 1.0$. Put values in italics indicate that immediate exercise is optimal.

Table – 3



Conclusion

It should now be clear that whenever stock price movements conform to a discrete binomial process, or to a limiting form of such a process, options can be priced solely on the basis of arbitrage considerations. Indeed, we could have significantly complicated the simple binomial process while still retaining this property.

The probabilities of an upward or downward move did not enter into the valuation formula. Hence, we would obtain the same result if *q* depended on the current or past stock prices or on other random variables. In addition, *u* and *d* could have been deterministic functions of time. More significantly, the size of the percentage changes in the stock price over each period could have depended on the stock price at the beginning of each period or on previous stock prices.[18] However, if the size of the changes were to depend on any other random variable, not itself perfectly correlated with the stock price then our argument will no longer hold. If any arbitrage result is then still possible, it will require the use of additional assets in the hedging portfolio.

We could also incorporate certain types of imperfections into the binomial option pricing approach, such as differential borrowing and lending rates and margin requirements. These can be shown to produce upper and lower bounds on option prices, outside of which risk less profitable arbitrage would be possible.

Since all existing preference-free option pricing results can be derived as limiting forms of a discrete two-state process, we might suspect that two-state stock price movements, with the qualifications mentioned above, must be in some sense necessary, as well as sufficient, to derive option pricing formulas based solely on arbitrage considerations. To price an option by arbitrage methods, there must exist a portfolio of other assets which exactly replicates in every state of nature the payoff received by an optimally exercised option. Our basic proposition is the following. Suppose, as we have, that markets are perfect, that changes in the interest rate are never random, and that changes in the stock price are always random. In a discrete time model, a necessary and sufficient condition for options of all maturities and striking prices to be priced by arbitrage using only the stock and bonds in the portfolio is that in each period.

(a) the stock price can change from its beginning-of-period value to only two ex-dividend values at the end of the period, and

(b) the dividends and the size of each of the two possible changes are presently known functions depending at most on: (i) current and past stock prices, (ii) current and past values of random variables whose changes in each period are perfectly correlated with the change in the stock price, and (iii) calendar time.

The sufficiency of the condition can be established by a straightforward application of the methods we have presented. Its necessity is implied by the discussion at the end of section 3.19.20.21

[18] Of course different option pricing formulas would result from these more complex stochastic processes. See Cox and Ross (1976) and Geske (1979). Nonetheless, all option pricing formulas in these papers can be derived as limiting forms of a properly specified discrete two- state process.

[19] Note that option values need not depend on the present stock price alone. In some cases. formal dependence on the entire series of past values of the stock price and other variables can be summarized in a small number of state variables.

[20] In some circumstances, it will be possible to value options by arbitrage when this condition does not hold by using additional assets in the hedging portfolio. The value of the option will then in general depend on the values of these other assets, although in certain cases only parameters describing their movement will be required.

[21] Merton's (1976) model, with both continuous and jump components. is a good example of a

This round out the principal conclusion of this paper: the simple two -state process is really the essential ingredient of option pricing by arbitrage methods. This is surprising perhaps given the mathematical complexities of some of the current models in this field. But it is reassuring to find such simple economic arguments at the heart of this powerful theory.

stock price process for which no exact option pricing formula is obtainable purely from arbitrage considerations. To obtain an exact formula it is necessary to impose .restrictions on the stochastic movements of other securities as Merton did or on investor preferences. For example Rubinstein 11976) has been able to derive the Black-Scholes option pricing formula under circumstances that do not admit arbitrage, by suitably restricting investor preferences. Additional problems arise when interest rates are stochastic, although Merton (1973) has shown that some arbitrage results may still be obtained.

References

Black. F. and M. Scholes. 1973, The pricing of options and corporate liabilities., Journal of

Political Economy J. 637 654.

Brennan. MJ. and E.S. Schwartz. 1977, The valuation of American put options. Journal of

Finance 32. 449 462..

Cox. J.e. and S.A. Ross. 1975. The pricing of options for jump processes. Rodney L. White

Center Working Paper no. 2-75 (University of Pennsylvania, Philadelphia. PA).

Cox. J.e. and S.A. Ross. 1976. The valuation of options for alternative stochastic proces..c;es.

Journal of Financial Economics 3.145.166. Geske. R., 1979, The valuation of compound options, Journal of Finahcial Economics 7. 63-8l. Uarrison, J.M. and D.M. Kreps. 1979. Martingales and arbitrage in multi period securities markets. Journal of Economic Theory 20. 381-408.

Merton. R.C. 1973. The theory of rational option pricing. Bell Journal of Economics and

Management Science 4. 141- 183.

Merton. R.C. 1976. Option pricing when underlying stock returns are discontimlous. Journal of

Financial Economics 3, 125-144.

Merton. R.e.. 1977. On the pricing of contingent claims and the Modigliani-Millet theorem.

Journal of Financial Economics 5, 241 -250. Parkinson. M.. 1977. Option pricing: The American put. Journal of Business SO. 21-36. Rendleman. RJ. and BJ. Bartler. 1978. Two-state option pricing, Unpublished paper (Graduate

School of Management. Northwestern University. Evanston. IL).

Rubinstein. M.. 1976. The valuation of uncertain income str.eams and the pricing of options. Bell

journal of Economics 7. 407 u425.

Sharpe. W.F.. 197R. Investments (Prentice-Hall. Englewood Cliffs. NJ).

ARBITRAGE PRICING THEORY

Objectives

- Understand of how APT relates to CAPM.
- Helps you understand the investors preference
- Analysis of both long term and short term securities.

Hello, you have already come across the CAPM and valuation of securities. Recall it just a minute.

The Capital Asset Pricing Model (CAPM) is an equilibrium model that de-scribes why different securities have different expected returns. In particular, this positive economic model *of* asset pricing asserts that securities have differ-ent expected returns because they have different betas. However, there exists an alternative model *of* asset pricing that was developed by Stephen Ross. It is known as Arbitrage Pricing Theory (APT), and in some ways it is less complicat-ed than the CAPM.

The CAPM requires a large number *of* assumptions, including those initially made by Harry Markowitz when he developed the basic mean-variance model. For example, each investor is assumed to choose his or her optimal portfolio by the use of indifference curves based on portfolio expected returns and standard deviations. In contrast, APT makes fewer assumptions. One primary APT as-sumption is that each investor, when given the opportunity to increase the re-turn of his or her portfolio without increasing its risk, will proceed to do so, the mechanism for doing so involves the use *of* arbitrage portfolios.

Factor Models

APT starts out by making the assumption that security returns are related to an unknown number of unknown factors.[1]For ease of exposition, imagine that there is only one factor and that factor is the predicted rate of increase in indus-trial production. In this situation, security returns are related to the following one-factor model:

$$r_i = a_i + b_i F_1 + e_i \quad (1)$$

where:

r_1 = rate of return on security

F_1 = the value of the factor which in this case is the predicted rate of growth in Industrial production

e_1 = random error term

In this equation, b_i is known as the **sensitivity** of security *i to* the factor. (It is also known as the factor loading for security *i* or the attribute of security *i.)* [2]

Imagine that an investor owns three stocks and the current market value of his or her holdings in each one is $4,000,000. In this case, the investor's current investable wealth W_0 is equal to $12,000,000. .Everyone believes that these three stocks have the, following expected returns and sensitivities:

i	r_1	b_1
Stock 1	15%	.9
Stock 2	21	3.0
Stock 3	12	1.8

Do these expected returns and factor sensitivities represent an equilibrium situation? If not, what will happen to stock prices and expected returns to re-store equilibrium?

Principle of Arbitrage

In. recent years, baseball card conventions have become commonplace events. Collectors gather to exchange baseball cards with one another at negotiated prices. Suppose that Ms. *A* attends such a. gathering where in one corner she finds *S* offering to sell a 1951 Mickey Mantle rookie card for $400. Exploring the convention further, she finds *B* bidding $500 for the same card. Recognizing a fi-nancial opportunity, Ms. *A* agrees to sell the card to *B*, who gives her $500 in cash. She races back to give $400 to S, receives the card, and returns with it to *B*, who takes possession of the card. Ms. *A* pockets the $100 in profit from the two transactions and moves on in search of other opportunities. Ms. A has engaged in a form of arbitrage.

Arbitrage is the earning of risk less profit by taking advantage of differential pricing for the same physical asset or security. As a widely applied investment tac-tic, arbitrage typically entails the sale of a security at a relatively high price and the simultaneous purchase of the same security (or its functional equivalent) at a relatively low price.

Arbitrage activity is a critical element of modern, efficient security markets. Because arbitrage profits are by definition risk less, all investors have an incentive to take advantage of them whenever they are discovered. Granted, some in-vestors have greater resources and inclination to engage in arbitrage than others. However, it takes relatively few of these active investors to exploit arbitrage situations' and, by their buying and selling actions, eliminate these profit opportunities.

The nature of arbitrage is clear when discussing different prices for an indi-vidual security. However, "almost arbitrage" opportu-nities can involve "similar"- securities or portfolios. That similarity can be defined in many ways. One inter-esting way is the exposure to pervasive factors that affect security prices.

A factor model implies that securities or portfolios with equal-factor sensitiv-ities will behave in the same way except for non-factor risk. Therefore, securities or portfolios with the same factor sensitivities should offer the same expected re-turns. If not, then "almost arbitrage" opportunities exist. Investors will take ad-vantage of these opportunities, causing their elimina-tion. That is the essential logic underlying APT. .

Arbitrage Portfolios

According to APT, an investor will explore the possibility of forming an **arbitrage portfolio** substantially to increase the expected return of his or her current port-folio without increasing its risk. Just what is an arbitrage portfolio? First of all, it is a portfolio that does not require any additional funds from the investor; If X_i denotes the *change* in the investor's holdings of security *i* (and hence the weight of security *i* in the arbitrage portfolio), this requirement of an arbitrage portfo-lio can be written as:

$$X_1 + X_2 + X_3 = 0. \quad (2)$$

Second, an arbitrage portfolio has no sensitivity to any factor. Because the sensi-tivity of a portfolio to a factor is just a weighted average of the sensitivities of the securities in, the portfolio to that factor, this requirement of an arbitrage portfo-lio can be written as:

$$b_1X_1 + b_2X_2 + b_3X_3 = 0 \quad (3a)$$

or, in the current example:

$$.9X_1 + 3.0X_2 + 1.8X_3 = 0. \quad (3b)$$

Thus, in this example, an arbitrage portfolio will have no sensitivity to industrial production.

Strictly speaking, an arbitrage portfolio should also have zero non-factor risk. However, the APT assumes that such risk is small enough to be ignored. In its terminology, an arbitrage portfolio has "zero factor exposures."

At this point many potential arbitrage portfolios can be identified. These, candidates are simply portfolios that meet the conditions given in Equations (2) and (3b). Note that there are three unknowns (X_1, X_2, and X_3) and two equations in this situation, which means that there is an infinite number of combinations of values for X_1, X_2, and X_3 that satisfy , these two equations.[3] As a way of finding one combination, consider arbitrarily assigning a value of.1 to X_1,

Doing so results in two equations and two unknowns:

$$.1+X_2+X_3=0 \quad (4a)$$

$$.09 + 3.0X_2 + 1.8X_3 = 0. \quad (4b)$$

The solution to Equations (4a) and (4b) is $X_2 = .075$ and $X_3 = -.175$. Hence a potential arbitrage portfolio is one with these weights.

In order to see if this candidate is indeed an arbitrage portfolio, its expected return must be determined. If it is positive, then an arbitrage portfolio will have been identified[4]. Mathematically, this third and last requirement for an arbitrage portfolio is:

$$X_1r_1 + X_2r_2 + X_3r_3 > 0 \quad (5a)$$

or, for this example,

$$15X_1 + 21X_2 + 12X_3 > 0. \quad (5b)$$

Using the solution for this candidate, it can be seen that its expected return is (15% X .1) + (21% X .075) + (12% X -.175) = .975%. Because this is a positive number, an arbitrage portfolio has indeed been identified.

The arbitrage portfolio just identified involves buying $1,200,000 of stock 1 and $900,000 of stock 2. How were these dollar figures arrived at? The solution comes from taking the current market value of the portfolio (W_0 = $12,000,000) and mul-tiplying it by the weights for the arbitrage portfolio of X_1 = .1 and X_2 = .075. Where does the money come from to make these purchases? It comes from selling $2,100,000 of stock 3. (Note that X_3W_0 = -.175 X $12,000,000 = -$2,100,000.)

In summary, this arbitrage portfolio is attractive to any investor who desires a higher, return and is not concerned with non-factor risk. It requires no additional dollar investment, it has no factor risk, and it has a positive expected return.

The Investors Position

At this juncture the investor can evaluate his or her position from either one of two equivalent viewpoints: (1) holding both the old portfolio and the arbitrage portfolio or (2) holding a new portfolio. Consider, for example, the weight in stock 1. The old portfolio weight was .33 and the arbitrage portfolio weight was .10, with the sum of these two weights being equal to .13. Note that the dollar value of the holdings of stock 1 in the new portfolio rises to $5,200,000 (= $4,000,000 + $1,200,000), so its weight is .43 (= $5,200,000/$12,000,000), equivalent to the sum of the old and arbitrage portfolio weights.

Similarly, the portfolio's expected return is equal to the sum of the expected returns of the old and arbitrage portfolios, or 16.975% (= 1.6% + .975%). Equivalently, the new portfolio's expected return can be calculated using the new portfolio's weights and the expected returns of the stocks, or 16.975% [= (.43 X 15%) + (.41 X 21 %) + (.16 X 12%)].

The sensitivity of the new portfolio is 1.9 [= (.43 X .9) + (41 X 3.0) + (.16 X 1.8)]. This is the same as the sum of the sensitivi-ties of the old and arbi-trage portfolios (= 1.9 + 0.0).

What about the risk of the new portfolio? Assume that the standard devia-tion of the old portfolio was 11 %. The variance of the arbitrage portfolio will be small because its only source of risk is non-factor risk. Similarly, the variance of the new portfolio will differ from that of the old only as a result of' changes in its non-factor risk. Thus it can be concluded chat the risk of the new portfolio will be approximately 11 %.[5] Tables 1 summarizes these observations.

	Old portfolio	+	Arbitrage Portfolio	New Portfolio
Weights:				
X_1	.333		.100	.433
X_2	.333		.075	.408
X_3	.333		-.175	.158
Properties:				
$\bar{r}_p$	1 6.000%		.975%	16.975%.
b_p	I. 900		.000	1.900
σ_p	11.000%		small	approx. 11.000%

Pricing Effects

What are the consequences of buying stocks 1 and 2 and selling stock 3? As everyone will be doing so, their market prices will be affected and, accordingly, their expected returns will adjust. Specifically, the prices of stocks 1 and 2 will rise because of increased buying pressure. In turn, this will cause their expected returns to fall. Conversely, the selling pressure put on stock 3 will cause its stock price to fall and its expected return to rise.

This can be seen by examining the equation for estimating a stock's expect-ed return:

$$\bar{r} = \frac{\bar{P}_1}{Po} - 1$$
(12.6)

where P_0 is the stock's current price and $\bar{P}_1$ is the stock's expected end-of-period price. Buying a stock such as stock 1 or 2 will push up its current price *Po* and thus result in a decline in its expected return *r*. Conversely, selling a stock such as stock 3 will push down its current price and result in a rise in its' expected return.

This buying-and-selling activity will continue until *all* arbitrage possibilities are significantly reduced or eliminated. At this point there will exist an approxi-mately linear relationship between expected returns arid sensitivities of the fol-lowing sort:

$\bar{r}_i$ = lo + l_1 b_i (7)

where A_o and A_1 are constants. This equation is the asset pricing equation of the APT when returns are generated by one factor.6 Note that it is the equation of a straight line, meaning that in equilibrium there will be a linear relationship be-tween expected returns and sensitivities.

In the example, one possible equilibrium setting could have A_0 = 8 and A1 = 4.[7] Consequently, the pricing equation is:

$\bar{r}_1 = 8 + 4b_i$. (8)

This would result in the following equilibrium levels of expected returns for stocks 1, 2, and 3:

$\bar{r}_1$ = 8 + (4 X .9) = 11.6%

$\bar{r}_2$ = 8 + (4 X 3.0) = 20.0%

$\bar{r}_3$ = 8 + (4 X 1.8) = 15.2%.

As a result, the expected returns for stocks 1 and 2 will have fallen from 15% and 21%, respectively, to 11.6% and 20% because of increased buying pressure. Con-versely, increased selling pressure will have caused the expected return on stock 3 to rise from 12% to 15.2%. The bottom line is that the expected return on al).y security is, in equilibrium, a linear function of the security's sensitivity to the factor b_i.

A Graphical Illustration

Figure 1 illustrates the asset pricing equation of Equation (7). Any security that has a factor sensitivity and expected return such that it lies off the line will be mispriced according to the APT and will present investors with the opportu-nity of forming arbitrage portfolios. Security *B* is an example. If an investor buys security *B* and sells security S in equal dollar amounts, then the investor will have formed an arbitrage portfolio.[8] How can this be?

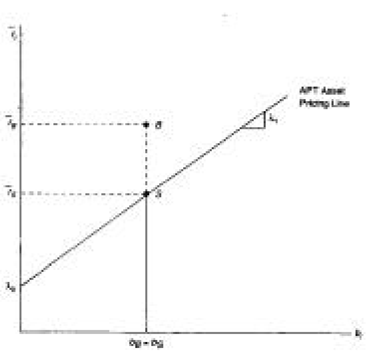


Figure 1

APT Asset Pricing Line

First of all, by selling an amount of security S to pay for the long position in security *B* the investor will not have committed any new funds. Second, because securities Band S have the same sensitivity to the factor, the selling of security S and buying of security *B* will constitute a portfolio with no sensitivity to the fac-tor. Finally, the arbitrage portfolio will have a positive expected return because the expected return of security *B* is greater than the expected return of security *S*.[9] *As* a result of investors buying security B, its price will rise and, in turn, its ex-pected return will fall until it is located on the APT asset pricing line.[l0]

Interpreting the APT Pricing Equation

How can the constants λ_0 and λ_1 that appear in the APT pricing Equation (12.7) be interpreted? Assuming that there is a riskfree asset in existence, such an asset will have a rate of return that is a constant. Therefore this asset will have no sen-sitivity to the factor. From Equation (12.7) it can be seen that $\lambda_1 = \lambda_0$ for any asset with $b_i = 0$. In the case of the riskfree asset, it is also known that $r_i = r_f$ im-plying that $\lambda_0 = r_f$. Hence the value of A_0 in Equation (12.7) must be r_f allow-ing this equation to be rewritten as

$r_i = r_f + \lambda_1 b_i$. (9)

In terms of λ_1, its value can be seen by considering a **pure factor portfolio** (or pure factor play) denoted *p** that has unit sensitivity to the factor, meaning $b_p^* = 1.0$. (If there were other factors, such a portfolio would be constructed so as to have no sensitivity to them.) According to Equation (12.9), such a portfolio will have the following expected return:

$r_{p*} = r_f + \lambda_1$
(12.10a)

Note that this equation can be rewritten as:

r_{p*} - rf = λ_1.
(12.10b)

Thus λ_1 is the expected excess return (meaning the expected return over and above the riskfree rate) on a portfolio that has unit sensitivity to the factor. Ac-cordingly, it is known as a factor risk premium (or factor-expected return premi-um). Letting $\delta_1 = r_{p*}$ denote the expected return on a portfolio that has unit sensitivity to the factor, Equation (l2.10b) can be rewritten as:

$\delta_1 - r_f = \lambda_1$.
(12.10c)

Inserting the left-hand side of Equation (10c) for λ_1 in Equa-tion (9) re-sults in a second version of the APT pricing equation:

$r_i = r_f + (\delta_1 - r_f)\, b_i$. (11)

In the example, *because* r_f = 8% and λ_1. = δ% - r_f = 4%, it follows that δ_1 = 12%. This means that the expected return on a portfolio that has unit sen-sitivity to the first factor is 12%.

In order to generalize the pricing equation of APT, the case where security returns are generated by more than one factor needs to be examined. This is done by considering a two-factor model next and then expanding the analysis to k factors where $k > 2$.

Two-factor Models

In the case of two factors, denoted F_1 and F_2 and representing predicted indus-trial production and inflation, each security will have two sensitivities, b_{i1} and b_{i2}. Thus security returns are generated by the following factor model:

$$r_i = a_i + b_{i1}F_1 + b_{i2}F_2 + e_i. \quad (12)$$

Consider a situation where there are four securities that have the following ex-pected returns and sensitivities.

i	r_1	b_{i1}	b_{i2}
Stock 1	15%	.9	2.0
Stock 2	21	3.0	1.5
Stock 3	12	1.8	.7
Stock 4	8	2.0	3.2

In addition, consider an investor who has $5,000,000 invested in each of the se-curities. (Thus the investor has initial wealth W_0 of $20,000,000.) Are these secu-rities priced in equilibrium?

12.3.1 Arbitrage Portfolios

To answer this question, the possibility of forming an arbitrage portfolio must be explored. First of all, an arbitrage portfolio must have weights that satisfy the fol-lowing equations:

$$X_1 + X_2 + X_3 + X_4 = 0 \quad (13)$$

$$.9X_1 + 3X_2 + 1.8X_3 + 2X_4 = 0 \quad (14)$$

$$2X_1 + 1.5X_2 + .7X_3 + 3.2X_4 = 0 \quad (15)$$

This means that the arbitrage portfolio must not involve an additional commit-ment of funds by the investor and must have zero sensitivity to each factor.

Note that there are three equations that need to be satisfied and that each .equation involves four unknowns. Because there are more unknowns than equa-tions, there are an infinite number of solutions. One solution can be found by setting X_1 equal to .1 (an arbitrarily chosen amount) and then solving for the re-maining weights. Doing so results in the following weights: $X_2 = .088$, $X_3 = -.108$, and $X_4 = -.08$.

These weights represent a potential arbitrage portfolio. What remains to be done is to see if this portfolio has a positive expected return. Calculating the ex-pected return of the portfolio reveals that it is equal to 1.41 % [= (.1 X 15%) + (.088 X 21%) + (-.108 X 12%) + (-.08 X 8%)]. Hence an arbitrage port-folio has been identified.

This arbitrage portfolio involves the purchase of stocks 1 and 2, funded by selling stocks 3 and 4. Consequently, the buying-and-selling pressures will drive the prices of stocks 1 and 2 up and stocks 3 and 4 down. In turn, this means that the expected returns of stocks 1 and 2 will fall and stocks 3 and 4 will rise. Investors will continue to create such arbitrage portfolios until equilibrium is reached. This means that equilibrium will be attained when any portfolio that satisfies the conditions given by Equations (13), (14), and (15) has an expected return of zero. This will occur when the following linear relationship between expected returns and sensitivities exists:

$$r_i = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2} \quad (16)$$

As in Equation (7), this is a linear equation except that it now has three dimensions, r_i, b_{i1} and b_{i2} Hence it corresponds to the equation of a two -dimensional plane.

In the example, one possible equilibrium setting is where $\lambda_0 = 8$, $\lambda_1 = 4$, and $\lambda 2. = -2$. Thus the pricing equation is:

$$r_i = 8 + 4b_{i1} - 2b_{i2} \quad (17)$$

As a result, the four stocks have the following equilibrium levels of expected returns:

r1 = 8 + (4 X .9) - (2 X 2) = 7.6%

r2 =.8 + (4 X 3) - (2 X 1.5) = 17.0%

r3 = 8 + (4 X 1.8) — (2 X .7) = 13.8%

r4 = 8 + (4 X 2) - {2 X 3.2) = 9.6%.

The expected returns of stocks 1 and 2 have fallen from 15% and 21 % while the expected returns of stocks 3 and 4 have risen from 12% and 8%, respectively. Given the buying-and-selling pressures generated by investing in arbitrage port-folios, these changes are in the predicted direction.

One consequence of Equation (17) is that a stock with higher sensitivity to the first factor than another stock will have a higher expected return if the two stocks also have the same sensitivity to the second factor because $ë_1 > O$. Conversely, since $\lambda_2 < 0$, a stock with higher sensitivity to the second factor will have a lower expected return than another stock with a lower sensitivity to the second factor, provided that both stocks have the same sensitivity to the first fac-tor. However, the effect of two stocks having different sensitivities to both factors can be confounding. For example, stock 4 has a lower expected return than stock 3 even though both of its sensitivities are larger. This is because the advan-tage of having a higher sensitivity to the first factor ($b_{41} = 2.0 > b_{31} = 1.8$) is not of sufficient magnitude to offset the disadvantage of having a higher sensitivity to the second factor ($b_{42} = 3.2 > b_{32} = .7$).

Pricing Effects

Extending the one-factor APT pricing Equation (7) to this two-factor situa-tion is relatively uncomplicated;-As before, ëo is equal to the riskfree rate. This is because the riskfree asset has no sensitivity to either factor, meaning that its val-ues of b_{i1} and b_{i2} are both zero. Hence it follows that $\lambda_0 .= r_f$ Thus Equation (16) can be rewritten more generally as:

$$r_i = rf + \lambda_1 b_{i1} + \lambda 2b_{i2} \quad (18)$$

In the example given in Equation (12.16), it can be seen that $r_f = 8\%$.

Next consider a well-diversified portfolio that has unit sensitiv-ity to the first factor and zero sensitivity to the second factor. *As* mentioned earlier, such a port-folio is known- as a pure factor portfolio or pure factor play because it has: (l) unit sensitivity to one factor, (2) no sensitivity to any other factor, and (3) zero non-factor risk. Specifically, it has $b_1 = l$ and $b_2 = 0$. It can be seen from Equa-tion (12.18) that the expected return on this portfolio, denoted δ_1, will be equal to $r_f + \lambda_1$. As it follows that $ä_1 - r_f = \lambda_1$. Equation (18) can be rewritten as:

$$r_L = r_f + (\delta_1 - r_f)\, b_{il} + \lambda_2 b_{i2} \qquad (19)$$

In the example given in Equation (12.16), it can be seen that $\delta_1 - r_f = 4$. This means that $\delta_1 = .12$ because $r_f = 8$. In other words, a portfolio that has unit sen-sitivity to predicted industrial production (the first factor) and zero sensitivity to predict inflation (the second factor) would have an expected return of 12%, or 4% more than the riskfree rate of 8%.

Finally, consider a portfolio that has zero sensitivity to the first factor and unit sensitivity to the second factor, meaning that it has $b_1 = 0$ and $b_2 = 1$. It can be seen from Equation (18) that the expected return on this portfolio, denot-ed ä$_2$, will be equal to $r_f + \lambda_2$. Accordingly, ä$_2 - r_f =$ ë$_2$ thereby allowing Equation (12.19) to be rewritten as:

$$r_i = r_f + (\delta_1 - rf)\, b_{il} + (\delta_2 - r_f)\, b_{i2}. \qquad (20)$$

In the example given in Equation (16), it can be seen that $\delta_2 - r_f = -2$. This means that ä$_2$ = 6 since $r_f = 8$. In other words, a portfolio that has zero sensitivity to predicted industrial production (the first factor) and unit sensitivity to predict inflation (the second factor) would have an expected return of 6% or 2% less than the risk free rate of 8%.

SINGLE FACTOR MODEL

Objective

- After completion of this lesson, you would be able to a means by which the investor can identify his or her optimal portfolio when there are an infinite number of possibilities.

Dear friends, if you really want to come to investment in a portfolio, the basic objective of modern portfolio theory is to provide a means by which the investor can identify his or her optimal portfolio when there are an infinite number of possibilities. Using a framework involving expected return and standard deviation for each security under consideration for inclusion in the portfolio along with all the covariance's between securities. With these estimates the investor can derive the curved efficient set of Markowiz. Then for a given riskfree rate the investor can identify the tangency portfolio and determine the location of the linear efficient set. Finally, the investor can proceed to invest in this tangency portfolio and borrow or lend at the riskfree rate, with the amount of borrowing or lending depending on the investor's risk-return preferences.

Factor Models And Retuen-generating Processes

The task of identifying the curved Markowitz efficient set can be greatly simplified by introducing a **return-generating process**. A return-generating process is a statistical model that describes how the return on a security is produced. Chapter 8 presented a type of return-generating process known as the market model. The market index. However there are many other type of return-generating processes for securities.

Factor Models

Factor models or index models assume that the return on a security is sensitive to the movements of various factors or indices. The market model assumes that to the there is one factor- the return on a market index. However, in attempting to accurately estimate expected returns, variances, and co variances for securities, multiple-factor models are potentially more useful than the market model. They have this potential because it appears that actual security returns are sensitive to more than movements in a market index. This means that there probably is more than one pervasive factor in the economy that affects security returns.

As a return-generating process, a factor model attempts to capture the major economic forces that systematically move the prices of all securities. Implicit in the construction of a factor model is the assumption that the returns on two securities will be correlated-that is, will move together- only through common reactions to one or more of the factors specified in the model. Any aspect of a security's return unexplained by the factor model is assumed to be unique or specific to the security and therefore uncorrelated with the unique elements of returns on other securities. As a result, a factor model is a powerful tool for portfolio management. It can supply the information needed to calculate expected returns, variances, and covariance's for every security- a necessary condition for determining the curved Markowitz efficient set. It can also be used to characterize a portfolio's sensitivity to movements in the factors.

Application

As a practical matter, all investors employ factor models whether they do so explicitly or implicitly. It is impossibly to consider separately the interrelationship or every security with every other. Numerically, the problem of calculating covariance's among securities rises exponentially as the number of securities analyzed increases.

Conceptually, thinking about the tangled web of security variances and covariance's become mind-boggling as the number of securities increases beyond just a few securities, let alone hundreds or thousand. Even the vast data processing capabilities of high-speed computers are strained when they are called upon to construct efficient sets from a large number of securities.

Abstraction is therefore an essential step in identifying the curved Markowiz efficient set. Factor models supply the necessary level of abstraction. They provide investment managers with a framework to identify important factor in the economy and the marketplace and to assess the extent to which different securities and portfolios will respond to changes in these factor.

Given the belief that one or more factors influence security returns, a primary goal of security analysis is to determine these factors and the sensitivities of security returns to movements in these factors. A formal statement of such a relationship is termed a factor model of security returns. The discussion begins with the simplest form of such a model, a one-factor model.

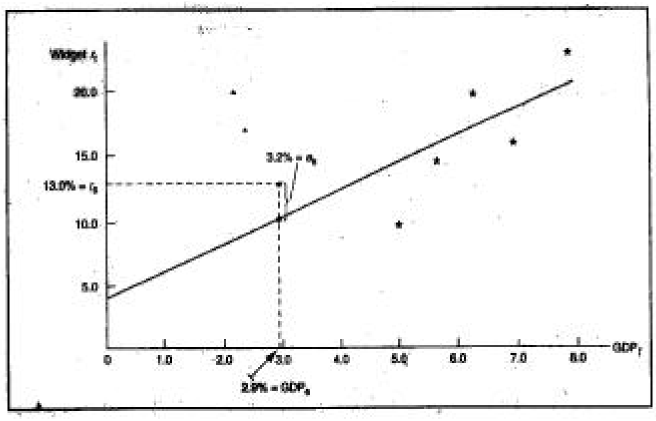
One-factor Models

Some investors argue that the return-generating process for securities involves a single factor. For example, they may contend that the reruns on securities on securities respond to the predicted growth rate in the gross domestic product (GDP). Table1 and Figure 1 illustrate one way of providing substance for such statements.

TABLE 1

FACTOR MODEL DATA

Year	Growth Rate in GDP	Rate of Inflation	Return on Widget Stock
1	5.7%	1.1%	14.3%
2	6.4	4.4	19.2
3	7.9	4.4	23.4
4	7.0	4.6	15.6
5	5.1	6.1	9.2
6	2.9	3.1	13.0



Two Important Features of One-Factor Models

Two features of one-factor models are of particular interest.

The Tangency Portfolio

First, the assumption that the returns on all securities respond to a single com-mon factor greatly simplifies the task of identifying the tangency portfolio. To determine the composi-tion of the tangency portfolio, the investor needs to esti-mate all the expected returns, variances, and co variances. This can be done with a one-factor model by estimating a_i' h_i, and s_{ei} for each of the N risky securities.[6]

Also needed are the expected value of the factor $\overline{F}$ and, its standard deviation, s_F. With these estimates, Equations (3), (4), and (5) can subsequently be used to calculate expected returns, variances, and. covariance's for the securities. Using these values, the curved efficient set of Markowitz can be derived. Finally, the tangency portfolio can be determined for a given risk free rate.

The common responsiveness of securities to the factor eliminates the need to estimate directly the co variances between the securities. Those co variances are captured by the securities' sensitivities to the factor and the factor's variance.

Diversification

The second interesting feature of one-factor models has to do with diversifica-tion. Earlier it was shown that diversification leads to an averaging of market risk and a reduction in unique risk. This feature is true of anyone-factor model ex-cept that instead of market and unique risk, the words factor and nonfactor risk are used. In Equation (4) the first term on the right-hand side ($b^2_i \sigma^2_F$) known as the factor risk of the security, and the second term (s^2_{ei}) is known as the non-factor (or idiosyncratic) risk of the security.

With a one-factor model, the variance of a portfolio is given by:

$$\sigma^2_P = b^2 = b^2_p \sigma F + \sigma^2_{ep} \qquad (6a)$$

Where

$$b_p = \sum_{i=1}^{N} X_i b_i \qquad (6b)$$

$$\sigma^2 ep = \sum_{i=1}^{N} X^2_i \sigma^2_{ei} \qquad (6c)$$

Equation (6a) shows that the total risk of any portfolio can be viewed as hav-ing two components similar to the two compo-nents of the total risk of an individ-ual security shown in Equation (4), In particular, the first and second terms on the right-hand side of Equation (6a) are the factor risk and non-factor risk of the portfolio, respectively.

As a portfolio becomes more diversified (meaning that it contains more se-curities), each proportion Xi will become smaller. However, this will not cause b_p to either decrease or increase significantly unless a deliberate attempt is made to do so by adding securities with values of bi that are either relatively low or high, respectively, As Equation (6b) shows, this is because b_p is simply a weighted av-erage of the sensitivities of the securities b_i with the values of Xi serving as the weights. Thus diversification leads to an averaging of factor risk.

However, as a portfolio becomes more diversified, there is **reason to expect** σ^2_{ep}, the non-factor risk, to decrease. This can be shown by examining Equation (6c).

Assuming that the amount invested in each security is equal, and then this equation can be rewritten by substituting $1/N$ for X_i:

$$\sigma^2 ep = \sum_{i=1}^{N} \left(\frac{1}{N}\right)^2 \sigma^2_{ei}$$

$$\left(\frac{1}{N}\right) \left(\frac{\sigma^2_{e1} + \sigma^2_{e2} + \ldots \sigma^2_{eN}}{N}\right)$$

The value inside the square brackets is the average non-factor risk for the individ-ual securities. But the portfolio's non-factor risk is only one-Nth as large as this because the term 1/ N appear outside the brackets. As the portfolio becomes more diversified, the number of securities in it, N, becomes larger. This means that 1/ N becomes smaller, which in turn reduces the non-factor risk of the port-folio. Simply stated, diversifica-tion reduces non-factor risk.[7]

MULTIPLE FACTOR MODEL

Objective

- After completion of this lesson, you would be able to a means by which the investor can identify his or her optimal portfolio when there are an infinite number of possibilities.

Any way you have read the lesson nineteen where we discussed about the single factor models. As a continuation we are coming to multiple factor models.

The health of the economy affects most firms. Thus changes in expectations con-cerning the future of the economy can be expected to have profound effects on the returns of most securities. However, the economy is not a simple, monolithic entity. Several common influences with pervasive effects might be identified.

1. The growth rate of gross domestic product
2. The level of interest rates
3. The inflation rate
4. The level of oil prices

Two-Factor Models

Instead of a one-factor model, a multiple-factor model for security returns that considers these various influences may be more accurate. As an example of a multiple-factor model, consider a two-factor model. This means assuming that the return-generating process contains two factors.

In equation form, the two-factor model for period *t* is:

$$r_{it} = a_i + b_{i1}F_{1t} + b_{i2}F_{2t} + e_{it} \quad 1)$$

where F_{1t} and F_{2t} are the two factors that are pervasive influences on security re-turns and b_{i1} and b_{i2} are the sensitivities of security *i* to these two factors. As with the one-factor model, e_{it} is a random error term and a_i is the expected return on security *i* if each factor has a value of zero.

Figure 1 provides an illustration of Widget Company's stock, whose re-turns are affected by expectations concerning both the growth rate in GDP and the rate of inflation. As was the case in the one-factor example, each point in the figure corresponds to a particular year. This time, however, each point is a com-bination of Widget's return, the rate of inflation, and the growth in GDP in that year as given in Table 1 at lesson no. 19. To this scatter of points is fit a two-dimensional plane by using the statistical technique of *multiple-regression analysis. (Multiple* refers to the fact that there are two predicted variables, GDP and inflation, in this exam-ple on the right-hand side of the equation.) The plane for a given security is de-scribed by the following adaptation of Equation (1):

$$r_t = a + b_1GDP_t + b_2INF_t + e_t.$$

The slope of the plane in the GDP growth-rate direction {the term b_1) repre-sents Widget's sensitivity to changes in GDP growth. The slope of the plane in the inflation rate direction (the term b_2) is Widget's sensitivity to changes in the inflation rate. Note that the sensitivities b_1 and b_2 in this example are positive

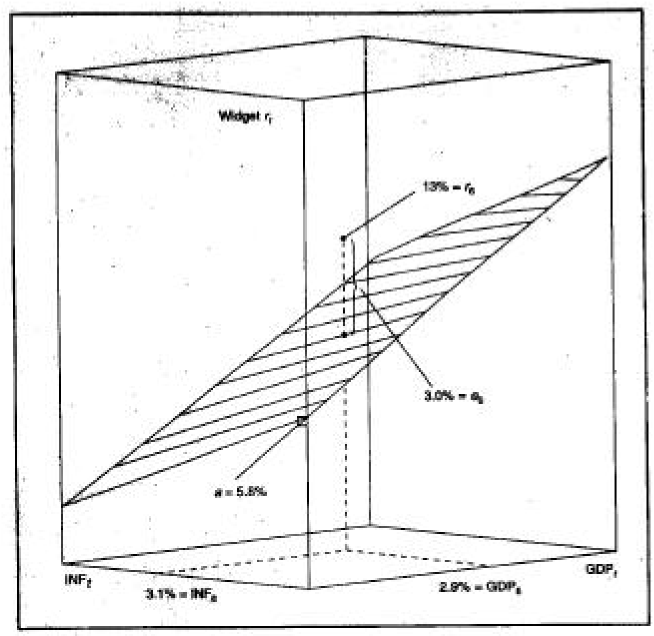


Figure 1

A Two-Factor Model

and negative, respectively, having corresponding values of 2.2 and - .7.[8] This in-dicates that as predicted GDP growth or inflation rises, Widget's return should increase or decrease, respectively.

The intercept term (the zero factor) in Figure 1 of 5.8% indicates Widget's expected return if both GDP growth and inflation are zero. Finally, in a given year the distance from Widget's actual point to the plane indicates its unique return (e_{it}), the portion of Widget's return not attributed to either GDP growth or infla-tion. For example, given that GDP grew by 2.9% and inflation was 3.1 %, Widget's expected return in year 6 equals 10% [= 5.8% + (2.2 X 2.9%) - (.7 X 3.1%)]. Hence its unique return for that year is equal to + 3% (= .13% - 10%).

Four parameters need to be estimated for each security with the two-factor model: a_i, b_{i1}, b_{i2}, and the standard deviation of the random error term denoted s_{ei}. For each of the factors, two parameters need to be estimated. These parame-ters are the expected value of each factor (F_1 and F_2) and the variance of each factor (s^2 and s^2). Finally, the covariance between the factors $COV(F_1F_2)$ needs to be estimated.

Expected Return

With these estimates, the expected return for any security *i* can be determined by using the following formula:

$$\overline{r} = a_i + b_{i1}\overline{F_1} + b_{i2}\overline{F}2 \quad (3)$$

For example, the expected return for Widget equals 8.9% [= 5.8% + (2.2 X 3%) - (.7 X 5%)] provided that the expected increases in GDP and inflation are 3% and 5%, respectively.

Variance

According to the two-factor model, the variance for any security is

$$s^2 = b^2 s^2_{F1} = b^2_{i2} s^2_{F2} + 2b_{i1} b_{i2} COV(F_1, F_2) + s^2 \quad (4)$$

If, in the example, the variances of the first (s^2_{F1}) and second (s^2) factors are equal to 3 and 2.9, respectively, and their covariance [COV(F_1, F_2)] equals .65, then the variance of Widget equals 32.1 [= (2.22 X 3) + (- .72 X 2.9) + (2 X 2.2 X-. 7 X .65) + 18.2], since its two sensitivities and random error term vari-ance are 2.2, - .7, arid 18.2, respectively.

Covariance

Similarly, according to the two-factor model the covariance between any two se-curities *i* and j can be determined by:

$$s_{ij} = b_{i1} b_{j1} s F^2_1 + b_{i2} b_{j2} s^2_2 + (b_{i1} b_{j2} + b_{i2} b_{j1}) \text{ COV } (F_1, F_2). \quad (5)$$

Thus continuing with the example, the covariance between Widget and Whatev-er is estimated to equal 39.9 { = (2.2 X 6 X 3) + (-.7 X -5 X 2.9) + [(2.2 X -5) + (-.7 X 6)] X .65} because the sensitivities of Whatever to the two factors are 6 and - 5, respectively.

The Tangency Portfolio

As with the one-factor model, once the expected returns, variances, and covari-ances have been determined using these .equations, the investor can proceed to use an *optimizer* (a special kind of mathematical routine) to derive the curved ef-ficient set of Markowitz. Then for a given riskfree rate, the tangency portfolio can be identified, after which the investor can deter-mine his or her optimal portfolio.

Diversification

Everything said earlier regarding one-factor models and the effects of diversifi-cation applies here as well.

1. Diversification leads to an averaging of factor risk.
2. Diversification can substantially reduce nonfactor risk.
3. For a well-diversified portfolio, nonfactor risk will be insignificant.

As with a one-factor model, the sensitivity of a portfolio to a particular factor in a multiple-factor model is a weighted average of the sensitivities of the securities where the weights are equal to the proportions invested in the securities. This can be seen by remembering that the return on a portfolio is a weighted average of the returns of its component securities

$$r_{pt} = \sum X_i r_{it} \quad (6)$$

Substituting the right-hand side of Equation (1) for r_{it} on the right-hand side of Equation (6) results in:

$$\begin{aligned} r_{pt} &= \sum_{i=1}^{N} X_i(a_i + b_{i1}F_{1t} + b_{i2}F_{2t} + e_{it}) \\ &= \left[\sum_{i=1}^{N} X_i a_i\right] + \left[\sum_{i=1}^{N} X_i b_{i1} F_{1t}\right] + \left[\sum_{i=1}^{N} X_i b_{i2} F_{2t}\right] + \left[\sum_{i=1}^{N} X_i e_{it}\right] \\ &= a_p + b_{p1}F_{1t} + b_{p2}F_{2t} + e_{pt} \end{aligned} \quad (11.12)$$

where:

$$a_p = \sum_{i=1}^{N} X_i a_i$$

$$b_{p1} = \sum_{i=1}^{N} X_i b_{i1}$$

$$b_{p2} = \sum_{i=1}^{N} X_i b_{i2}$$

$$e_{pt} = \sum_{i=1}^{N} X_i e_{it}.$$

Here, figure 11.12 given above should be referring to as figure 7 for your purpose.

Note that the portfolio sensitivities b_{p1} and b_{p2} are weighted averages of the respective individual sensitivities b_{i1} and b_{i2}

Sector-Factor Models

The prices of securities in the same industry or economic sector often move to-gether in response to changes in prospects for that sector. Some investor's ac-knowledge this by using a special kind of multiple-factor model referred to as a sector-factor model. To create a sector-factor model, each security under consid-eration must be assigned to a sector. For a two-sector-factor model, there are two- sectors and each security must be assigned to one of them.

For example, let sector-factor 1 consist of all industrial compa-nies and sector- factor 2 consist of all no industrial companies (such as utility, transportation, and financial companies). Thus F_1 and F_2 can be thought of as representing the re-turns on an industrial stock index and a non-industrial stock respectively. (They could, for example, be components of the S&P 500.) It should be kept in mind, however, that both the number of sectors and what each sector consists of is an open matter that is left to the investor to decide.

With this two-sector-factor model, the return-generating process for securi-ties is of the same general form as the two-factor model given in Equation (11.7). However, with the two-sector-factor model, *F}* and *F2* now denote sector-factors 1 and 2, respectively. Furthermore, any particular security belongs to either sector factor 1 or sector-factor 2 but not both. By definition, a value of zero is given to the sensitivity term corresponding to the sector-factor to which the security is not assigned. This means that either b_{i1} or b_{i2} is set equal to zero, depending on the sector-factor to which security i is not assigned. The value of the other sensi-tivity term must be estimated. (To make matters simple, some people simply give it a value of one.).

As an illustration, consider General Motors (GM) and Delta Air Lines (DAL). The two-sector-factor model for GM (the time subscript *t* has been delet-ed for ease of exposition here) would be:

$$r_{GM} = a_{GM} + b_{GM\,1}F_1 + b_{GM\,2}F_2 + e_{GM}. \quad (8)$$

However, because GM belongs to sector-factor 1 as an industrial security, the co-efficient b_{GM2} is assigned a value of zero. Once this assignment is made, Equa-tion (11.13) reduces to:

$$r_{GM} = a_{GM} + b_{GM1} F_1 + e_{GM}. \quad (9)$$

Thus only the values of a_{GM}, b_{GM1}, and O_{eGM} need to be estimated for GM with the two-sector-factor model. In comparison with the two-factor model, values of a_{GM}, b_{GM1}, b_{GM2}, and O_{eGM} need to be estimated.

Similarly, as DAL belongs to the nonindustrial sector, it would have the fol-lowing two-sector-factor model:

$$r_{DAL} = a_{DAL} + b_{DAL1} F_1 + b_{DAL2} F_2 + e_{DAL} \quad (10)$$

Which would reduce to:

$$r_{DAL} = a_{DAL} + b_{DAL2} F_2 + e_{DAL} \quad (11)$$

as b_{DAL1} would be assigned a value of zero. Thus only the values of a_{DAL}, b_{DAL2}, and O_{eDAL} need to be estimated with the two-sector-factor model.

In general, whereas four parameters need to be estimated for each security with a two-factor model (a_i, b_{i1}, b_{i2}, and O_{ei}), only three parameters need to be estimated with a two-sector-factor model (a_i, O_{ei}, and either b_{i1} or b_{i2}). With these estimates in hand, along with estimates of F_1, F_2, O_{F1}, and O_{F2}, the investor can use Equations (11.8) and (11.9) to estimate expected returns and variances for each security. Pairwise covariances can be estimated using Equation (11.10). This will then enable the investor to derive the curved efficient set of Markowitz from which the tangency portfolio can be determined for a given riskfree rate.

Extending the Model

To extend the discussion to more than two factors requires the abandonment of diagrams as the analysis moves beyond three dimensions. Nevertheless, the concepts are the same. If there are *k* factors, the multiple-factor model can be written as:

$$r_{it} = a_i + b_{i1} F_{1t} + b_{i2} F_{2t} + \ldots + b_{ik} F_{kt} + e_{it} \quad (12)$$

Where each security has *k* sensitivities, one for each of the *k* factors.

It is possible to have both factors and sector factors represented in Equation (12). For example, F_1 and F_2 could represent GDP and inflation as in Table 1in lesson 19, whereas F_3 and F_4 could represent the returns on industrial stocks and non-industrial stocks, respectively. Hence each stock would have three sensitivi-ties: b_{i1}, b_{i2}, and b_{i3} for industrials and b_{i1}, b_{i2}, and b_{i4} for no industrials.

ESTIMATING FACTOR MODELS

Objective

- After completion of this lesson, you would be able to a means by which the investor can identify his or her optimal portfolio when there are an infinite number of possibilities.

Let's continue the remaining part of the previous lesson. This is the-ESTIMATING FACTOR MODELS.

Although many methods of estimating factor models are used, these methods can be grouped into three primary approaches:

1. Time-series approaches
2. Cross-sectional approaches
3. Factor-analytic approaches

Time-series Approaches

Time-series approaches are perhaps the most intuitive to investors. The model builder with the assumption that he or she knows advance the factors that influence security returns. Identification of the relevant factors typically proceeds from an economic analysis of the firms involved. Aspects of macroeconomics, microeconomic, industrial organization, and fundamental security analysis will play a major role in process.

For example, as discussed earlier, certain macroeconomic variables might be expected to have a pervasive impact on security returns, including such things as predicted growth in GDP, inflation, interest rates, and oil prices. With these factors specified, the model builder collects information concerning the values of the factors and security returns from period to period. Using this data, the model builder can calculate the sensitivities of the securities' returns to the factors, the securities' zero factors and unique returns, and the standard deviations of the factors and their correlations. In this approach, accurate measurement of factor values is crucial. In practice, this can be quite difficult.

The Importance of Expectation

Security prices reflect investors' estimates of the present values of firms' future prospects. At any given time the price of Widget stock id likely to depend on the projected growth rate of GDP, the projected rate of inflation, and other factors.

Of investors' projections of such fundamental economic conditions change, so too will the price of Widget. Because the return on a stock is influenced heavily by changes in its price, stock returns are expected to be mire highly correlated with changes in expected future values of fundamental economic variables than with the actual changes that occur contemporaneously.

For example, a large increase in inflation that was fully anticipated might have no effect of the stock price of a company whose earnings are highly sensitive to inflation. However, of the consensus an expectation was for a low inflation rate, and then the subsequent large increase would have a large effect on the company's stock price.

For the reason, whenever possible if is desirable to select factors that measure changes in expectations rather than realizations, as the latter typically include both changes that were anticipated and those that were not. One way to accomplish this goal is to rely on variables that involve changes in market prices. Thus the difference in the returns on two portfolios—one consisting of stocks thought to be unaffected by inflation—can be used as a factor that measures revisions in inflation expectations. Those who construct factor models the time- series approach often rely on market- based surrogates for changes in forecasts of fundamental economic variables in this manner.

An Example

Table 1 and Figure 1in the previous lesson 20 presented an example of how to use the time-series approach to estimate a factor model. In this example, returns on individual stocks such as Widget were related to two factors—gross domestic product and inflation—by comparing over time each stock's returns to the predicted values of the factors.

Recently, Fama and French conducted a study that used a time-series approach to identify the factors that explain stock and bond returns.[12] In their study, monthly stock returns were found to be related to three factors: a market factor, and a book-to – market equity factor. In equation form, their factor model for appears as:

$$r_{it} - r_{ft} = a_i + b_{i1}(r_{Mt} - r_{ft}) + b_{i2}\,\text{SMB}_t + b_{i3}\text{HML}_t + e_{it}. \qquad (1)$$

The first factor $(r_{Mt} - r_{ft})$ is simple the monthly return on a broad stock market over and above the return on one-month Treasury bills on a broad stock index. The size factor (SMB_t) can be thought of as the difference in the monthly return on two stock indices—a small stock index and a big-stock index. (Here a stock's size is measured by its stock price at the end of June each year times the number of shares it has outstanding at that time. The small-stock index consists of stocks that are below the median NYSE size and the big-stock index consists of stocks that are above the median.) The book-to-market equity factor (HML_t) is also the difference on the monthly return on two stock indices—an index of stocks with high book-to-market equity is stockholders' equity taken from the firm's balance sheet and market equity is the same as the stock's size used in determining the previous factor. The high ratio index consists of stocks that are in the top third, and the low ratio index consists of stocks that are in bottom third.)

Fama and French also identified two factors that seem to explain monthly bond returns. In equation form, their factor model for bonds appears as:

$$r_{it} - r_{ft} = a_i + b_{i1}\text{TERM}_t + b_{i2}\text{DEF}_t + e_{it}. \qquad (2)$$

These tow factors are a term-structure factor and a default factor. The term structure factor (TERM_t) is simply the difference in the monthly returns on long-term Treasury bonds and one-month Treasury bills. The default factor (DEF_t) is the difference in the monthly returns on a portfolio of long-term corporate bonds and long-term Treasury bonds.

Cross-Sectional Approaches

Cross-sectional approaches are less intuitive than time-series approaches but can often be just as powerful a tool. The model builder begins with estimates of securities' sensitivities to certain factors. The in a particular time period, the values of the factors are estimated based on securities' returns and their sensitivities to the factors. This process is repeated over multiple time periods, thereby providing an estimate of the factors' standard deviations and their correlations.

Note that the cross-sectional approach is entirely different from the time-series approach. With the latter approach, the values of the factors are known and the sensitivities are estimated. Furthermore, the analysis is conducted for one security over multiple time periods, then another security, then another, and so on. With the former approach the sensitivities are known and the values of the factors are estimated. Accordingly, the sensitivities in the cross-sectional approach are sometime referred to as attribute. Furthermore, the analysis is conducted over one time period for a group of securities, then another time period for the same group, then another, and so on. Examples of one-factor and two-factor models will be shown next to illustrate the cross-sectional approach.

One-Factor Models

Figure 1 provides a hypothetical example of the relationship between the returns for a number of different stocks in a given time period and one security attribute—dividend yield—for each stock. Each point represents one particular stock, showing its return and dividend yield for the time period under evaluation. In this case, stocks with higher-dividend yields tended to do better—that is, have higher returns—than those with lower-dividend yields. Whereas figure 1 (an example of the time-series approach) is based on many stocks for one time period is based on one stock for many time periods.

To quantify the relationship shown in Figure 1, a straight line has been fitted to the diagram by using the statistical technique of simple-regression analysis. The equation of the line in Figure 1 is:

$$\bar{r}_{it} = 4 + .5b_{it} \qquad (3)$$

or, more generally

$$\bar{r}_{it} = a_t + b_{it}F_t \qquad (4)$$

where

$\bar{r}_{it}$ = *the expected return on stock i in period* t, *given that the factor had on actual value of* F_t,

a_t = *the zero factor in period* t,

b_{it} = *the dividend yield of stock I in period* t,

F_t = *the actual value of the factor in period* t.

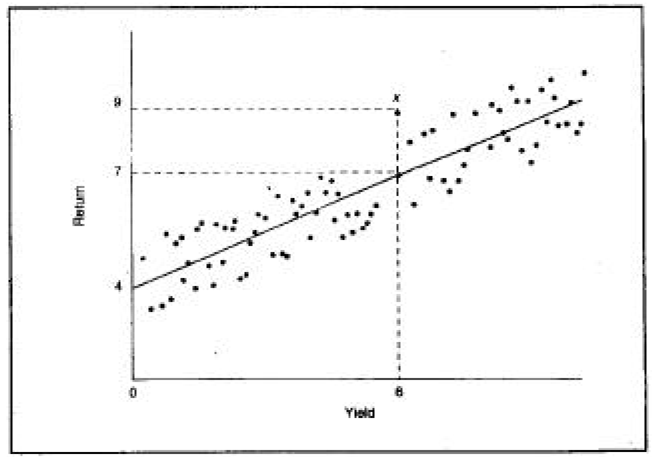


Figure 1

A Cross-Sectional One-Factor Model

The vertical intercept at indicates the expected return on a typical stock with a dividend yield of zero. In Figure 1 it is equal to 4%. The slope of .5 indicates the increase in expected return for each percent of dividend yield. Hence it represents the actual value of the dividend yield factor (F_t) in this time period.

From this example it can be seen that the cross-sectional approach uses sensitivities to provide estimates of the values of the factors. Hence these factors are known as empirical factors, In comparison it was shown earlier that the tike-series approach uses known values of factors to provide estimates of a security's sensitivities. Hence these factors are known as fun *mental factors.*

The actual return on any given security may lie above or below the line due to its nonfactor return. A complete description of the relationship for this one-factor model is:

$$r_{it} = 4 + .5b_{it} + e_{it} \qquad (5)$$

where e_{it} denotes the nonfactor return during period *t* on security *i*. In Figure1, security *x* had a dividend yield of 6%. Hence from Equation (11.18) it had an expected return during this time period of 7% [= 4 + (.5 X 6)]. Because it actually had a return of 9%, its nonfactor return was + 2% = 9% - 7%.

In periods such as the one shown in Figure 1, high-yield stocks tended to outperform low-yield stocks. This indicates that the yield factor F_t was positive at this time. However, in another time period it is possible that low-yield stocks will tend to outperform high-yield stocks. The regression line in the corresponding diagram will be downward-sloping, and the yield factor will be negative. In still other time periods, there will be no relationship between yield and return, resulting in a flat regression line and a yield factor of zero.

A Two-Factor Example

In some time periods small stocks tend to outperform large stocks. In other months the converse is true. Hence many cross-sectional models use a size attribute that is often computed by taking the logarithm of the total market value of the firm's outstanding equity measured in millions, which is, in turn, calculated by taking the firm's stock price and multiplying it by the number of shares outstanding equity measured in millions, which is, in turn, calculated by taking the firm's stock price and multiplying it by the number of shares outstanding and then dividing the resulting figure by one million. Thus a $1 million stock would be assigned a size attribute value of zero; a $10-million stock and so on. This convention is based on the empirical observation that he impact of the size factor on a security with a large total market value is likely to be twice as great as that on a security with one-tenth the value. More succinctly, the size effect appears to be *linear in the logarithms.*

To estimate the size factor in a given month, the procedure used in Figure 1 to estimate the yield factor can be employed. The size attributes of securities can be plotted on the horizontal axis and their returns for the given time period plotted (as in Figure 1) on the vertical axis. The slope of the resultant regression line provides an estimate of the size factor for the time period.

This procedure has drawbacks, however. Large stocks tend to have high yields. Thus differences in returns between large and small stocks may be due to some extent to differences in yield, not size. The estimated size factor may be in part a reflection of a true yield factor. The problem is symmetrical in that the estimated yield factor may also be in part a reflection of the true size factor.

Multiple- regression analysis is typically used to tit a plane to the data. In the example shown in Figure 11.4 this results on the following regression equation:

$$r_{it} = 7 + 4b_{i1t} - .3b_{i2t} + e_{it} \qquad (6)$$

Where b_{ilt} and b_{i2t} denote, respectively, the dividend yield and size of stock *i* in time period *t.* In general, the regression equation for a two-factor model is:

$$r_{it} = a_t + b_{ilt}F_{lt} + b_{i2t}F_{2t} + e_{it} \qquad (7)$$

Where at denotes the zero factor in time period *t* and the two factors are denoted F_{lt} and $F_{2t.}$

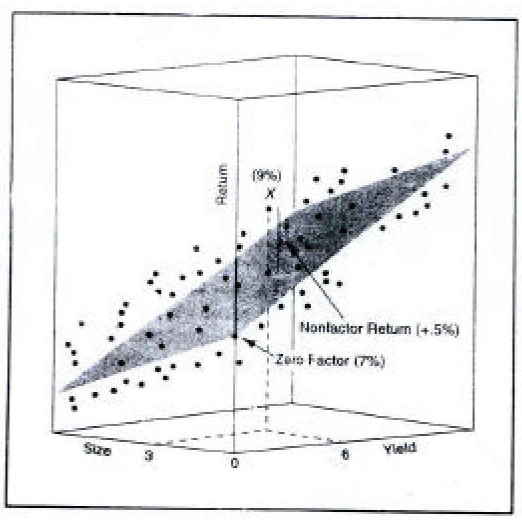


Figure 2

A Cross-section Two-Factor Model

The equation of the plane shown in Figure 11.4 is:

$$r_{it} = 7 + .4b_{ilt} - .3b_{i2t} \qquad (8)$$

or, more generally,

$$\overline{r}_{it} = a_t = b_{ilt}F_{lt} + b_{i2t}F_{2t} \qquad (9)$$

This means that the zero factor a_t was 7%, including that a stock with zero dividend yield and zero size (meaning a market value of $1 million) would have been expected to have a return of 7%. Note that the estimated values of the dividend yield-factor (F_{lt}) and size factor (F_{2t}) are .4 and -.3, respectively. Thus during this time period, higher-dividend yields and smaller sizes were both associated with larger returns.

Using Equations (7) and (9) a given security x with a dividend yield of 6% and a size of 3 would have been expected to have a return of 8.5% [= 7 + (.4 X 60 – (.3 X 3)]. With an actual return of 9% its nonfactor return e_{it} is thus +.5% (= 9% - 8.5%) during this time period, as shown in Figure 2.

The inclusion of size and dividend yield along with the use of multiple- regression analysis can help sort it the effects of differences in yield and size on differences in security returns. It cannot deal adequately with influences that ate not represented at all, nor can it guarantee that the included attributes are not simply serving as *proxies* for other, more fundamental at-tributes. Statistical tests can indicate the ability of the variables included in the analysis to explain or predict past security returns. But judgment and luck are required to identify variables that can help predict future security returns, risk, and covari-ances. The extension to more than two variables follows in a straightforward manner from what has been indicated in Equations (11.23) through (11.26).

An Example

Sharpe conducted a study that used a cross-sectional approach to identify the factors that explain stock returns.[17] In his study, stock returns were related monthly to five security sensitivities (and eight sector-factors) that were measured for each stock. These sensitivities consisted of firm size (measured like Fama and French), the stock's historic beta when measured against a stock market index, the stock's dividend yield, the stock's historic beta when measured against a bond market index, and a measure of how much of the stock's historical return wan due to mis-pricing.

Over the period 1931 to 1979, Sharpe collected monthly data for over 2,000 common stocks. Among other results, his analysis produced a value of .237 (annualized) for the dividend yield factor. By implication then, a stock with a 5% dividend yield outperformed a stock with a 4% dividend yield (but with similar exposures to all the other factors) by almost.24% per year.

Factor-Analytic Approaches

Finally, with factor-analytic approaches the model builder knows neither the factor values nor the securities' sensitivities to those factors. A statistical technique called *factor analysis* is used to extract the number of factors and securities' sensitivities based simply on a set of securities' past returns. Factor analysis takes the returns over many time periods on a sample of securities and attempts to identify one or more statistically significant facture that could have generated the co variances of returns observed within the sample. In essence, the return data tells the models the models builders about the structure of the factors models. Unfortunately, factors analysis does not specify what economic variables the factors represent.

Limitations

There is no reason to assume that a food factor model for one period will be a good one for the next period. Key factors change as in the effect of energy prices on security markets in the 1970s and more recently during the war in the Persian Gulf. The risk and returns associated with the relevant factors nor their magnitudes could be applied to security returns over an extended past period and the sensitivities of securities to factors can change over time.

It would be convenient if neither the relevant factors nor their magnitudes were to change from period to period. If this were so, mechanical procedures could be applied to security returns over an extended past period and the factor model inferred along with all the needed magnitudes. As it is, statistical estimation methods should be tempered with the judgment of the model builder to account for the dynamic nature of the investment environment.

Questions and Problems

1. Included among the factors that might be expected to be pervasive are expectations regarding growth in real GNP, real interest rate, inflation, and oil prices. For each factor, provided an example of an industry i.e. expected to have a high (either positive or negative sensitivity) to the factor.
2. Why do factor models greatly simplify the process of deriving the curved Markowitz efficient set?
3. Many investment management firms assign each of their security analysts to research a particular group pf stocks. (Usually these assignments are organized by industry.) How are these assignments an implicit recognition of the validity of factor-model relationships?
4. What are two critical assumptions underlying any factor model? Cite hypothetical examples of violations of those assumptions.
5. Cupid Childs, a wise investment statistician, once said with respect to factor models, “Similar stocks should display similar returns.” What did Cupid mean by this statement?
6. Based on a one-factor model, consider a security with a zero-factor value of 4% and a sensitivity to the factor of .50. The factor takes on a value of 10%. The security generates a return of 11%. What portion is related to nonfactor elements?
7. Based on a one-factor model, consider a portfolio of two securities with the following characteristics:

	Factor	Nonfactor	
Security	**Sensitivity**	**Risk( )**	**Proportion**
A	.20	49	40
B	3.50	100	.60

a. If the standard deviation of the factor is 15%, what is the factor risk of the portfolio?

b. What is the nonfactor risk of the portfolio?

c. What is the portfolio’s standard deviation?

8. Recalculate the answer to problem 7 assuming that the portfolio is also invested in the riskfree asset so that its investment proportions are

Security	Proportion
Riskfree	.10
A	.36
B	.54

9. Based on a one-factor model, security *A* has a sensitivity of -.50, whereas security *B* has a sensitivity of 1.25. If the covariance between the two securities is –312.50, what is the standard deviation of the factor?
10. Based on a one-factor model, for two securities *A* and *B*:

    $r_A = 5\% + .8F + e_A$

    $r_B = 7\% + 1.2F + e_B$

    $\sigma_F = 18\%$

    $\sigma_{eA} = 25\%$

    $\sigma_{eB} = 15\%$

    Calculate the standard deviation of each security.
11. Based on a one-factor model, if the average nonfactor risk (σ^2_{ei}) of all securities is 225, what is the nonfactor risk of a portfolio with equal weights assigned to its 10 securities? 100 securities? 1,000 securities?
12. Based on the discussion of factor and nonfactor risk and given a set of securities that can be combined into various portfolios, what might be a useful measure of the relative diversification of each of the alternative portfolios?
13. With a five-factor model (assuming uncorrelated factors and a 30-stock portfolio, how many parameters must be estimated to calculate the expected return and standard deviation of the portfolio? How many additional parameter estimates are required if the factor are correlated?
14. Based on a three-factor model, consider a portfolio composed of three securities with the following characteristics:

Security	Factor1 Sensitivity	Factor2 Sensitivity	Factor3 Sensitivity	Proportion
A	-.20	3.60	05	60
B	.50	10.00	75	20
C	1.50	2.20	30	20

15. Dode Cicero owns a portfolio of tow securities. Based on a two-factor model, the two securities have the following characteristics:

Security proportion	Zero Factor	Factor1 sensitivity	Factor2 sensitivity	Non-Factor risk	(s^2_{ei})
A	2%	30	2.0	196	.70
B	3	50	1.8	100	.30

The factors are uncorrelated. Factor 1 has an expected value of 15% and a standard deviation of 20%. Factor 2 has as expected value of 4% and a standard deviation of 5%. Calculate the expected return and standard deviation of Dode's portfolio. [Hint: Think about how Equation (11.6a) could be extended to a two-factor model by considering Equation (11.9).]

16. Compare and contrast the three approaches to estimating factor models.
17. Consider a factor model with earning yield (or earnings/ price ratio) and book-price (or book value/ market price ratio) as the two factors. Stock *A* has an earnings yield of 10% and a book-price of 2. Stock B's earnings yield is 15% and it's book-price is 90. The zero factors of stocks *A* and B are 7% and 9%, respectively. Of the expected returns of stocks *A* and *B* are 18% and 16.5%, respectively, what are the expected earnings-yield and book-price factor values?
18. Based on a two-factor model, consider two securities with the following characteristics:

Characteristic	Security A	Security B
Factor 1 sensitivity	1.5	.7
Factor 2 sensitivity	2.6	1.72
Nonfactor risk (s^2_{ei})	25.0	16.0

The standard deviations if factor 1 and factor 2 are 20% and 15%, respectively and the factors have a covariance of 225. What are the standard deviations of securities *A* and *B*? What us their covariance?

LESSON 22:
TUTORIALS

Banc One Corporation: Asset and Liability Management

On November 15, 1993, Dick Lodge, Banc One Corporation's (Banc One's) chief in-vestment officer (CIO), gathered his notes and headed for a meeting with John B. McCoy, Banc One's chairman and CEO. On the way, he recalled the lunchtime conversation on the golf course six weeks earlier, during which McCoy had first voiced concern over Banc One's falling share price-from a high of $483/4 in April 1993 to just $36 3/4 (see Exhibit 1). McCoy attributed the decline to investor concern over Banc One's large and growing interest rate derivatives portfolio. During their discussion in September, McCoy had asked Lodge, who was responsible for managing the bank's investment and derivatives portfolio, to think about ways to deal with this problem.

McCoy had been prompted into action not only by the continued price decline, but also by the com-ments of equity analysts who covered Bane One:

The increased use of interest rate swaps is creat-ing some sizable distortions in reported earnings, reported earning assets, margins, and the historical measure of return on assets. . . Were Bane One to include [swaps] in reported earning assets, the ad-justed level would be 26% higher than is currently reported. . . Given its large position in swaps, Bane One overstates its margin by 131% [and its! return on assets in excess of 0.20%. . . Adjusted for [swaps], Bane One's tangible equity-to-asset ratio would de-cline by 1.55%.[3]

Banc One's investors are uncomfortable with so much deriva-tives exposure. Buyers of regional banks do not expect heavy derivatives involvement. . . Heavy swaps usage clouds Bane One's- financial image [and is] extremely confusing. . . It is virtually impossible for anyone on the outside to assess the risks being assumed. [4].

What made this situation more perplexing was that Banc One already had attempted to pre-empt concern over its growing derivatives portfolio. Along with its second-quarter results, it distributed a book-let detailing its asset and liability manage-ment poli-cies and describing its derivatives portfolio, which had grown during the quarter from $23.4 billion to $31.5 billion in notional principal[5]. Lodge and others believed that the information in the booklet would help assuage any investor's concerns. Yet, given these kinds of comments from the analysts, the message was clearly not getting through.

In Lodge's mind, there was a simple explanation for the large size of Banc One's derivatives portfolio: swaps were attractive investments that lowered the bank's exposure to movements in interest rates. Why the market was penalizing Banc One for something that *reduced* its exposure to risk remained a mystery to him. Earlier in the year, Lodge had expressed his puzzlemen(to a reporter: "Why in the world more banks don't look at interest rate swaps. . . I don't know. It's not an esoteric phenomenon anymore. [6]

Nevertheless, he knew that *McCoy* attributed the decline to the derivatives portfolio and wanted to discuss alternatives for dealing with the situation.

Banc One Corporation[7]

Banc One Corporation, headquartered in Co-lumbus, Ohio, truly epitomized the spirit of re-gional banking. With $76.5 billion in assets, it was the largest bank holding company based in Ohio and the eighth largest in the country. Unlike the more traditional bank holding company structure, in which the parent corporation controlled subsid-iary banks, Banc One had a three-tiered organiza-tional structure operating across 12 states. The parent, Banc One Corporation, controlled 5 state bank holding companies (in Arizona, Indiana, Ohio, Texas, and Wisconsin), which in turn owned. 42 subsidiary banks, or "affiliates." Through its Re-gional Affiliate Group, Banc One owned another 36 subsidiary banks-for a total of 78 banking affiliates. In addition to its banking affiliates, Banc One controlled 10 nonbanking organizations in various businesses ranging from insurance to venture capital to data processing. .

For its banking business, Banc One had a very well defined, three-pronged strategy: concentrate on retail and middle-market commercial customers; use technology to enhance customer service and to assist in the management of banking affiliates; and grow rapidly by acquiring profitable banks.

Since 1969, it had completed 76 acquisitions involving 139 banks. In just the 10 years since 1982, it had completed 50 acquisitions, making it one of the top 10 corporate acquirers in the country.8 As of November 1993; Banc One had ten pending acqui-sitions that would bring an additional S9 billion in assets to the corporation. One of the largest pending acquisitions was Liberty National Bancorp, a bank holding company in Louisville, Kentucky with $4.7 billion in assets.

This deal highlighted many of the principles that guided Banc One's acquisitions. The target, Liberty National, had a strong retail focus, had a solid management team, and was the market leader. In addition, the deal was structured like most of its previous acquisitions: it would be accounted for as a: pooling of interests, be paid for with stock, and consist of a tiered offer that depended on the value of Banc One's stock price. The terms of the Liberty National Bancorp deal were as follows:

Banc One's Stock Price	Ratio of Banc one's Shares to Liberty National's Shares
Under $41.57	0.8421
$ 41.57 to $ 44.00	$ 35.00 worth of stock
above $ 44.00	0.7954

As of mid-November, Banc One's stock was trading near the "walkaway" price of $34.55. If it was below $34.55 in the second quarter of 1994, when the deal was expected to be consummated, one of two things would happen. Either Liberty National would cancel the deal or Banc One would end up using stock that it felt was undervalued to pay for the deal. Thus, a low stock price would either bring Banc One's acquisition program to a halt or cause it to violate one of its cardinal rules of acquisitions: acquisitions should not be dilutive. According to John McCoy, Banc One has "very strong pricing discipline. We just don't do dilutive acquisitions".[9] William Boardman, an Executive Vice President at Banc One, elaborated: "When we talk to prospects, we tell them we want the deal. to be non-dilutive when we do it, but that we also want it to be non--dilutive next year, and the year after that. Basically, what that means is that you have to grow your earnings at the same rate we're [Banc One] growing our earnings."[10]

While a strict set of principles guided Banc One's acquisition strategy, another well-defined set of principles guided its operating strategy. Internally, the operating strategy was known as the "uncom-mon partnership," which described the relationship among the affiliate banks and the various parts of the corporation. According to this partnership, the cor-poration decentralized the "people" side of the business and centralized the "paper" side. To capture the local knowledge of customers and markets, Banc One retained existing management in acquisitions and gave affiliate managers complete autonomy in running their banks. In contrast, Banc One central-ized all of the affiliates' data processing, record keeping, and back office operations. This centraliza-tion fit well with Banc One's growth strategy. According to Boardman, "Growing just to become larger is not part of our strategy. Growing our economies of scale is."[11] The centralization of opera-tions also capitalized on Banc One's vast experience with computer systems.

Over the years, Banc One had invested heavily in technology and information systems to support the uncommon partnership. Starting at the top with John B. McCoy, there was the belief that informa-tion was critical to running such a decentralized organization. One of the most important jobs of Banc One was to gather information from and dis-seminate it to the affiliates using the Management Information and Control System (MICS). This data-base tracked financial, productivity, and perfor-mance data for all affiliates. Every month, affiliates entered into the database their results and their re-vised budgets. In return, all affiliate presidents re-ceived a one-inch-thick report containing compara-tive statistics ranking all affiliates. The objective of this system was to encourage friendly competition among banking affiliates and to encourage managers to share information about effective banking products and practices.

Although it was an extremely complicated and highly decentralized organization, Banc One had one of the best financial track records of any bank in the country. Compared with the financial perfor-mance of the country's 25 largest bank holding companies in the decade since 1982, it had the highest average return on assets, the highest average return on equity, and the highest ratio of common equity to assets. Even more incredible was that Banc One had a string of 24 years of increasing earnings per share; none of the other large banks had a string of more than 7.[12]

Exhibit 2 summarizes Banc One's operating results and financial performance during the period 1983 to the third quarter of 1993.

Asset and Liability Management

A typical U.S. bank's liabilities consisted of floating-rate liabilities (such as federal funds borrow-ings) and long-term fixed-rate liabilities (such as certificates of deposit, or CDs). Assets included floating-rate assets (such as variable-rate mortgages and loans, as well as floating-rate investments) and long-term fixed-rate assets (such as fixed-rate mort-gages and securities). Asset and liability manage-ment involved matching the economic characteris-tics of a bank's inflows and outflows. For example, a bank could match the maturity of its assets and liabilities. It also could look at the duration, the contractual fixed/floating nature of its commitments, or an estimate of the period in which its commit-ments would be repriced in response to changes in market rates as the basis upon which to judge just how well it was matched.

Banks' needs to match assets to liabilities arose from their strategic decisions regarding interest rate exposure. A bank could engineer its assets and li-abilities to ensure that its earnings or market value would be unaffected by changes in interest .rates. Alternatively, a bank could adjust its portfolio of assets and liabilities to profit when rates rose, but lose when they fell. It could also position itself to gain when rates fell, and lose when they rose. The selec-tion of interest rate exposure was a major policy decision for financial institutions.

In practice, banks typically had relative

In practice, banks typically had relatively more long-term fixed-rate liabilities (such as CDs) than they had long-term fixed-rate assets (such as loans). To make up for this shortfall, banks that wished to match assets and liabilities complemented their loan portfolios with fixed-rate investments commonly called balancing assets, such as Treasury securities. By adjusting the characteristics of the balancing assets, a bank could better match its assets to its existing liabilities.

As chief investment officer of Banc One, Dick Lodge managed the firm's portfolio of balancing assets. His staff of approximately 100 people, with 12 engaged in asset and liability management activities, measured the degree to which the bank's assets and liabilities were matched and made profitable investments consistent with the bank's policy of managing its interest rate exposure. Specifically, they had an official mandate to (1) invest funds in conventional investments and derivatives to conserve the funds' principal value yet provide a reasonable rate of return; (2) keep enough funds in liquid investments to allow the bank to react quickly to demands for cash; (3) control the exposure of Banc One's re-ported earnings to swings in interest rates; and (4) achieve these objectives without unnecessarily increasing the bank's capital requirements. [13]

In carrying out this mandate, Banc One used investments and derivatives as substitutes for one another. For example, if it wanted to increase its share of fixed-return investments, it could sell a floating-rate investment (or borrow at a floating rate) and use the proceeds to buy a three-year fixed-rate Treasury note. The initial net outflow of these two transactions would be zero, but the transactions would increase the relative magnitude of the bank's fixed-rate portfolio. Alternatively, Banc One could enter into an interest rate swap in which it paid a floating rate of interest and received a fixed rate in return. The initial net outflow of such a swap also would be zero. As in the first example, such a transaction would increase the bank's fixed-rate inflows and reduce its periodic net floating-rate inflows. Because the security transactions and the

swap produced similar interest rate exposure, they -had to be compared on other dimensions, such as yield, credit risk, capital requirements, transaction costs, and liquidity.

Defining and Measuring Interest Rate Exposure

Banc One, like other banks, defined its expo-sure to interest rate risk by calculating its earnings sensitivity, or the impact of interest rate changes on reported earnings. For example, if a gradual 1% upward shift in interest rates during the year in-creased that year's base earnings by 5%, the bank would have an earnings sensitivity of 5%. If earn-ings sensitivity was positive, the bank was said to be *asset sensitive* (i.e., the interest rate on assets reset more quickly than liabilities, resulting in in-creased income if rates rose). If earnings sensitivity was negative, the bank was said to be *liability sensitive* (i.e., liabilities reset more quickly than assets, resulting in a decrease in income if rates rose). If the bank had a 0% earnings sensitivity,

then an upward or downward shift in interest rates would have no effect on its earnings.

Like many banks, Banc One's basic portfolio (excluding its balancing assets) was asset sensitive. Its asset sensitivity arose because a large proportion of its assets, such as commercial loans, were indexed to the prime rate and therefore varied contractually with market rates. However, the bank's liabilities included mostly fixed-rate items such as fixed-rate CDs as well as "sticky-fixed" savings and demand deposits whose rates changed much more slowly than market indices. Banc One's relative overabun-dance of fixed-rate liabilities would make its earnings increase as rates rose. This natural asset-sensi-tivity was exacerbated by its acquisition program because many of the banks it acquired were highly asset sensitive.

Over the years, Banc One's evolving program to measure interest rate risk mirrored best practice in the U.S. banking industry. Prior to the 1980s, the bank did not precisely measure its exposure to changes in interest rates. Instead, it generally avoided investing in longer-maturity securities, feeling that these investments could add undue risk to the liquidity of its investment portfolio. By the early 1980s, it had become clear to Banc One's manage-ment that measuring interest rate risk was a critical task. The second oil shock of the 1970s had increased the level and volatility of interest rates. For example, the prime rate soared to more than 20% in late 1980, twice the average for the 1970s and four times as large as the average in the 1960s. In 1980 alone, the prime rose to 19.8% in April, fell to 11.1% in August and rebounded to more than 20% at the close of the year. To determine the bank's exposure to interest rate move-ments in this new, more volatile interest rate environment, Banc One began measuring its maturity gap in 1981.

Maturity gap analysis compared the difference in maturity between assets and liabilities, adjusted for their repricing interval. Repricing interval re-ferred to the amount of time over which the interest rate on an individual contract remained fixed. For example, a three-year loan with a rate reset after year one would have a repricing-adjusted maturity of one year. Banc One grouped its assets and liabilities into categories, or "buckets," on the basis of their repric-ing-adjusted maturities (less than 3 months, 3 to 6 months, 6 to 12 months, and more than 12 months). The maturity gap for each category was the dollar value of assets less liabilities. If the bank made short- term floating-rate loans funded by long-term fixed- rate deposits, it would have a large positive maturity gap in the shorter catego-ries and a large negative maturity gap in the longer periods.

The maturity gaps could then be used to predict how the bank's net interest margin (the difference between the weighted average interest rate received on assets and the weighted average interest rate paid on liabilities)-and therefore earnings-would be affected by changes in interest rates. For example, if interest rates dropped sharply, a large positive maturity gap for the short maturity buckets would predict a drop in interest income and. therefore earnings, because the bank would immediately receive lower rates on its loans while still paying higher fixed rates on its deposits.

Unfortunately, implementing the initial matu-rity gap measure-ment program was extremely time consuming. By the time each gap report was col-lected from the affiliates, consolidated, and ana-lyzed, the information was dated. Lodge himself con-structed the first gap management report in 1981, and it took almost a year to complete.

In 1984, Banc One began using asset and liability simulations as a more accurate method measure its exposure to interest rates. By using exact asset and liability portfolios rather than grouping each asset or liability according to its repricing interval, Banc One was able to measure how interest rate changes would affect earnings. To do so, it created an "on-line balance sheet" that contained up-to-date information on its assets and liabilities, which complemented the MICS process. The key features of each contract, including principal amounts, inter-est rates, maturity dates, and any amortization sched-ules of assets and liabilities, were recorded. Then, Banc One used historical data to estimate such items as the maturity of demand-deposit (checking) ac-counts, the speed with which its bank managers would reprice deposits and loans in response to interest rate shifts, and the rate at which its borrowers might refinance fixed-rate loans if rates dropped..

Once the model was complete, Banc One could simulate how any shift in interest rates would affect its balance sheet and earnings, as well as run sensitivity analyses on its assumptions. Although the model had been refined since 1984, it served as the basis for measuring the bank's interest rate risk and senior management reviewed its predictions monthly. In 1993, this on-line balance sheet was redesigned to include a monthly down-load of each of over 3 million loans or deposits, that is, a discrete asset and liability database on each cus-tomer that included prepayment, optionality, and convexity estimates.[14]

Investments for Managing Interest Rate Exposure

Banc One's evolving sophistication in manag-ing interest rate exposure mirrored its sophistication in measuring it. In the early 1980s, it managed its exposure to interest rate risk by adding balancing assets to its investment portfolio until it felt it had enough fixed-rate investments to offset its fixed-rate liabilities. In 1981, 13% and 21% pf Banc One's earning assets were money market investments and longer-term securities, respectively. Initially, Banc One invested in short- and medium-term U.S. Trea-suries and high-quality municipal bonds. Municipal bonds were an especially attractive investment be-cause prior to 1986, banks could deduct 800% of the interest expense incurred on monies raised to buy them. Because the income earned on the bonds was free of state and federal taxes, banks could enjoy a large after-tax spread on their leveraged municipal bond investments.

In 1983, Banc One began using interest rate swaps as part of its investment portfolio. Originally, swaps were used to lock in high after-tax yields on municipal securities. By buying the municipal bonds, Banc One .received an after-tax yield of 9.50%. By then entering into an interest rate swap in which it paid a fixed rate of 7.00% and received the London Interbank Offered Rate (LIBOR), a commonly used floating-rate index, it ended up with a net position of receiving LIBOR + *2.50010.* The bank's net cash flow *from* the investment and swap resembled a floating -rate investment with an above-market yield. During the course of 1983 and 1984, Banc One became increasingly comfortable with the use of swaps as a tool to tailor individual investments to suit its needs.

In 1986, Congress passed the Tax Reform Act, which eliminated *for* banks the deduction of interest expense on the financing *for* municipal bond invest-ments.[15] Banks turned to other invest-ments that would provide the same high yield they had grown accustomed to receiving. Banc One replaced many of its municipal investments with mortgage-backed securities (MBSs), which were fixed-income invest-ments whose payment stream was backed by pools of mortgage loans and which were typically guaran-teed by the federal government. MBSs provided a slightly lower promised after-tax yield than did municipal bonds and carried an additional risk of prepayment. If interest rates *fell,* borrowers typically refinanced their mortgages by prepaying their exist-ing mortgages. The owner of a pool of mortgages was forced to reinvest precisely when market yields were relatively low and was left with a submarket yield when rates *rose.*

In 1983, Wall Street created anew type of mortgage security: the CMO, or collateralized mort-gage obligation. CMOs took a pool of mortgage loans and carved the principal and interest outflows into a set of different securities, or trenches. The trenches differed from one another only in their priority for repayments of principal. For example, the first tranche of a CMO would receive all of the mortgage prepayments until its principal was re-turned to its holders. At that point, the second tranche would begin to receive prepayments until its principal was fully paid out, and so on. With a large pool of mortgages, investors could statistically esti-mate the likely speed of prepayment and therefore the likely time at which each tranche would be fully paid down and stop paying interest. Each tranche paid a different yield to compensate *for* the various amount of prepayment risk a buyer faced, as well as *for* the different average life of the investments. By investing in CMOs, Banc One could still receive the high yields associated with mortgage securities, assuming it was comfortable with the prepayment risk it would bear. In 1993, Banc One had 54.5 billion invested in CMOs, or about a third of their investment portfolio. Earlier in the 1980s, as much as two-thirds of their investment portfolio was held in CMOs.

Swaps as Synthetic Investments

After using swaps in the mid-1980s to tailor cash flows of individual municipal investments, Bane One realized that it could also use swaps as a proxy *for* some of its conventional fixed-rate investments. Instead of investing in medium-term *U.S.* Treasury obligations, it could simply enter into a medium-term receive-fixed swap and put its money into short-term floating-rate cash equivalents. There were several advantages of this "synthetic investment" over con-ventional investments.

First, the swap greatly improved the bank's liquidity. Banks need cash to accommodate cus-tomer withdrawals and to repay existing liabilities, such as CDs, as they mature. Investing in long-dated securities could increase a bank's yield, but if the bank needed to raise cash suddenly, these invest-ments might not be easily liquidated or their liquida-tion might expose the bank to a large loss in principal. With a swap, the bank could invest in short-term, highly liquid securities with stable prin-cipal values. By layering a receive-fixed swap onto this investment, the bank could obtain the econom-ics of the longer-term investment, while still enjoying the high liquidity of the short-term instrument.

Second, unlike investments and borrowings, swaps were off-balance-sheet transactions. If Bane One were to buy a fixed-rate bond and sell a floating- rate security, both would appear on its balance sheet, and the spread between the two would appear as income. However, if it were to enter into a receive- fixed swap with the same cash flow implications, the swap would not appear as either an asset or liability, but would be disclosed only in footnotes to the financial statements. Yet the current net income or loss from the swap transaction still would appear on its income statement. This accounting treatment would tend to overstate traditional profitability mea-sures such as a bank's return on assets in comparison to the identical securities transactions.

Finally, in comparison to a conventional secu-rities investment, swaps could also reduce the amount of capital needed to meet regulatory requirements. These minimum capital requirements grew out of an international agreement, the Basle Accord, signed by the central bankers of the major industrialized coun-tries. In agreement with the Accord, U.S. banking regulators implemented risk-based capital standards beginning in Decem-ber 1990. The new regulations dictated the amount of capital banks needed to hold as a function of their total risk-based assets.[16] As of year-end 1992, u.s. regulators raised the mini-mum capital levels and strengthened their power to close institutions that failed to meet these minimums.

Stricter capital standards led banks to prefer assets with lower capital requirements, all else being equal. Some observers attributed the rising growth in bank investments in Treasury securities to their zero risk weighting in the calculation of risk-adjusted assets. Under the capital guidelines, swaps contributed little to the risk-adjusted assets against which the bank had to hold capital.[17] Were a bank to create exposure similar to the swap using securities (other than U.S. Treasury securities), its need to hold capital would be *20010* to 100% of the principal value of the assets.[18]

During the late 1980s, Banc One began replac-ing many of its maturing conven-tional investments with synthetic investments. As part of this trend, it began to investigate whether it could create a synthetic CMO, which would have the advantages of other swaps, yet deliver the risk/return charac-teristics of CMO invest-ments. Specifically, a syn-thetic CMO would allow Banc One to enjoy high yields in exchange for taking on prepayment risk. After a few false starts and discussions with various investment banks, Banc One and its counterparties developed a product called Amortizing Interest Rate Swaps (AIRS).[19]

Because AIRS replicated investments in mort-gage securities, they needed to have similar prepay-ment features. With low interest rates, consumers prepay their mortgages, and mortgage investors receive back their principal. In the AIRS, the notional amount of the swap would be reduced or amortized if interest rates fell. As interest rates declined, the AIRS would amortize faster, thereby leaving the bank to reinvest just when market yields were low. Likewise, when interest rates increased, the maturity of an AIRS would end up longer than expected, thereby leaving the bank with a below-market yield on its investment. In early AIRS, the amortization of the notional principal balance was tied to the performance of a particular pool of actual mort-gages, but with later AIRS, the amortization schedule was set by a formula. Exhibit 3, panel A, gives the terms for the latter type of AIRS.

As synthetic investments, AIRS produced attractive yields. In these transactions, Banc One would receive a fixed rate of interest and pay LIBOR. In 1993, this fixed rate, called a swap spread, was perhaps 120 basis points over a Treasury security of the same maturity. In comparison, the bank could buy a comparable CMO and receive a yield of 100 basis points over Treasuries. If Banc One was to enter into a standard (non-amortizing) swap of the same term, it might receive a fixed rate of 20 basis points over Treasuries.

With Banc One's mortgage portfolio as well as its investments in CMOs and AIRS, prepayment risk complicated the task of measuring interest rate risk. The embedded options that Banc One sold to its mortgage borrowers, certain depositors, and to its swap counterparties made its earnings sensitivity nonlinear. With a rise in rates, the earnings from its fixed-rate investments would not change. However, a drop in rates which precipitated prepayments of mortgages or amortization of the AIRS forced the

EXHIBIT 3

REPRESENTATIVE SWAP

TRANSACTIONS

PANEL A : AMORTIZING INTEREST RATE SWAP (AIRS) SEPTEMBER 1993

Notional amount ………….. $ 500 million.

Final maturity……………… 3 years (if not amortized early)

Payment Frequency……….. Quarterly

Banc one pays……………….3 – month LIBOR (3.25% at initiation of swap)

Banc one receives ………….. 4.5%

Lock out period …………….1 year

(During the lockout period, there is no amortization of swap)

Cleanup provision ………… 10% of original national amount
(If the notional amount falls to $50 million or less through amortization, the swap is cancelled)

Amortization schedule…….. Each quarter, after the lockout period, the notional principal of the swap is reduced by the following amount for the following quarter, depending on the level of interest rates.

If 3 – month LIBOR	Notional Principal Amount	Average Life of swap
Stays at 3.35% or falls	Completely amortized	1.25 years
Rises to 4.35%	Reduced by 31%	1.75 years
Rises to 5.35%	Reduced by 10.5%	2.50 years
Rises to 6.35 or higher	Not reduced	3.25 years

PANEL B: LIBOR – PRIME BASIS SWAP

Notional amount ……………… $ 200 million

Final maturity ………………… 4 years

Payment frequency …………… Quarterly

Banc One pays ………………. Daily average prime rate – 270 basis points
(At initiation, prime was 6%)

Banc One receives …………… 3-month LIBOR (subject to caps)
(At initiation, 3 month LIBOR was 3.375%)

Caps …………………………… In no quarterly period can the rate Banc One receives exceed 25 basis point over the rate received in the prior quarter.

bank to reinvest the early repayment of principal at the lower market rates.. Furthermore, steep rate drops typically increased the rates of prepayment or amortization. For example, though earnings might drop 1% for a 1% increase in rates, a 2% increase in rates might reduce earnings by 3% or 4%, not 2%.

Swaps as a Tool for Risk Management

Banc One had a long-standing stated policy of “minimizing the impact of fluctuating interest rates on earnings and market values,”2O and in 1986, its senior management adopted guidelines for allow-able earnings sensitivity. This first policy stated that earnings could not change more than 5% for a 1% immediate change in interest rates. Because Banc One was more asset sensitive than its policy would permit, the bank considered alternatives for adjust-ing its earnings sensitivity, finally using swaps as its solution.

Although in the past the bank had entered into pay-fixed swaps to transform the cash flows on its municipal investments, the exact opposite swap was required to shift it away from an asset-sensitive position and toward more liability sensitivity. By entering into an interest rate swap in which it paid a floating rate and received a fixed rate in return, it was as if the bank was incurring a floating-rate liability while investing in a fixed-rate asset. This combination would move the bank toward a liabil-ity-sensitive (or negative earnings-sensitive) posi-tion. If interest rates rose, the floating-rate payments on the swaps would increase the bank’s interest expense while interest income remained constant, thus reducing earnings and producing liability sen-sitivity. As Banc One gradually enlarged its interest rate swap portfolio in the mid-1980s, its earnings sensitivity moved to within the specified 5% bound-ary. See Exhibit 4 for historical information on Banc One’s investment portfolio, swap portfolio, maturity gap, and earnings sensitivity during the period 1988 through the third quarter of 1993.

EXHIBIT 4 BANC ONES INVESTMENT PORTFOLIO AND INTEREST RATE SENSITIVITY / 1998 THROUGH 1995 Q3($ MILLIONS)

	Investments							Swaps			
	Amount Outstanding				Gross Income Received			Gross	Maturity		
	Earnings Assets	Loans	Short-term Investments	Securities	Loans	Short term Investment	Securities	Amount Outstanding	Income Received.[b]	Gap Measures[c]	Earnings Sensitivity
1988	$22,531	$17,325	$581	$4,625	$1,876	$28	$368	N/A	N/A	-6.67%	-1.00%
1989	23,568	17,909	525	5,133	2,167	39	446	$3,299	$291	-3.59%	-1.00%
1990	26,680	20,363	628	5,272	2,303	58	441	3,231	292	-10.07%	-1.55%
1991	41,482	31,168	2,324	7,989	2,747	61	484	11,214	887	-7.33%	-2.30%
1992	54,766	39,142	1,740	13,884	3,872	86	870	10,492	766	-15.70%	-2.61%
1993:Q1	61,807	45,361	1,382	15,064	1,159	11	231	14,132	240	-2.34%	-2.50%
1993:Q2	66,796	48,845	1,978	15,973	1,173	9	235	17,280	275	-2.65%	-2.60%
1993:Q3	68,116	50,105	1,217	16,794	1,189	10	216	22,515	335	-3.64%	-3.30%

a. Includes only receive-fixed swaps.

b. Notional volume of outstanding receive-fixed swaps multiplied by average fixed rate received on such swaps.

c. Maturity gap over the first one-year horizon as a percentage of earning assets, where maturity gap is defined as total assets with adjusted maturity of one year or less minus total liabilities with adjusted maturity of one year or less.

Sources: Banc One Corporation. Annual Reports. 10-Ks.

Because the swaps were designed to adjust the bank’s *earnings* sensitivity, the greater its earnings, the more swaps it would need. Also, the more its natural earnings sensitivity strayed from the policy guidelines, the more swaps it would need. Both of these factors contributed to the subsequent growth in its swap portfolio. For example, in 1989, Banc One’ acquired banks with $12 billion in assets from M Corp, a failed Texas bank. These banks were 23.4% asset sensitive when they were bought, far outside Banc One’s policy target range and well above its then-slight liability sensitivity. To bring the new banks in line, Lodge had to enter into a large notional volume of swaps. The bank’s continued acquisition strategy, as well as its earnings growth, would increase its need for swaps.[21]

Managing Basis Risk

Though synthetic investments reduced Banc. One’s earnings sensitivity to overall shifts in interest rates, they created a heightened sensitivity to mis-matches *between* floating-rate interest rates, or basis risk. Most of Banc One’s floating-rate assets were based on the prime rate. However, most conven-tional interest rate swaps as well as its AIRS used three-month LIBOR as an index for floating-rate payments. LIBOR was an actively traded global market rate that changed daily. In contrast, the prime rate was an administered U.S. or local rate that changed infrequently at bankers' discretion. Because of these differences, the spread between the two rates changed dramatically over time. (See Exhibit 5 for a graph of prime and three-month LIBOR.)

For example, assume the baI1k entered into a swap in which it received 7% and paid LIBOR. Ignoring the difference between prime and LIBOR, it would effectively transform its prime-based float-ing-rate assets into fixed-rate investments paying 7%. However, if three-month LIBOR increased 150 basis points but prime was unchanged, Banc One would have transformed its prime-based floating-rate asset into a fixed-rate asset paying not 7% but 5.5%, and it would have created basis risk through its exposure to swings in the prime-LIBOR spread.

To counter this basis risk, Banc One entered into basis swaps that reduced the floating-rate mis-match (see Exhibit 3, panel B, for typical basis swap terms), In a basis swap, Banc One would pay a floating rate based on the prime rate and receive a floating rate based on three-month LIBOR. This contract would offset the spread differential between prime and three-month LIBOR. Using a basis swap in conjunction with an AIRS in which it paid LIBOR, Banc One could confidently transform prime-based floating-rate assets to fixed-rate investments.

Managing Counterparty Risk

The credit risk of investing in swaps differed from that of traditional investments, If Banc One bought a U.S. Treasury bond, for example, it would face no credit risk. However, if it entered into a swap transaction in which it received the fixed rate, it would be exposed to the default of its counterparty.

This credit risk was mitigated in three ways. First, the positive swap spread (i.e. "yields on swaps were higher than on Treasury securities") gave the

EXHIBIT 6

BANC ONE'S EXPOSURE TO ITS MAJOR SWAP COUNTERPARTIES OCTOBER 31, 1993

($ IN MILLIONS)

	National Amount	Average Maturity	Mark to Market Exposure[a]	Collateral Posted[b]	Net MTM Exposure[c]	Potential Exposure[d]	Net Credit Exposure[e]
Bankers Trust	$12,142	1.77	$123	$132	($9)	$ 68	$ 59
Union Bank of Switzerland	6,976	1.87	49	49	0	92	92
Goldman Sachs	6,163	1.57	58	122	(64)	26	(38)
Lehman Brothers	4,058	2.32	16	81	(65)	26	(39)
Merrill Lynch	3,347	2.17	59	104	(45)	10	(35)

a. Mark to market exposure measured as the market value of swap positions with counterparty, A positive exposure indicates that Bane One's swaps have a market value greater than zero.

b. Collateral is posted in the form of cash or bank-eligible securities. A positive number indicates that Bane One's counterparties have deposited collateral with Bane One.

c. Represent. mark to marker (MTM) exposure less collateral posted by Bane One's counterparties.

d. The bank estimated its potential exposure if it experienced a large movement in interest rates relative to historical experience. Specifically using historic data, it calculated the distribution of interest rate moves Over 30 days. It then calculated how much it could lose. If rates moved in Bane One's favor, and if the size of the rare move was equal to a three-standard deviation change in rates. 99% of all rate moves would he within three standard deviations. So this measurement of its potential gains was considered a conservative estimate of the bank's credit exposure.

e. Represent. Bane One's potential loss less the collateral it currently has on hand. Source: Bane One Corporation.bank a higher return to compensate for its credit risk. Second, in an investment, the bank's entire principal was at risk (if the issuer was not the u.s. govern-ment), whereas in a swap, only the net payment (fixed less floating) was at risk of default. Third, Banc One established strict policies for managing its counterparty exposure.

In all instances, its counterparties were rated no lower than single-A. To understand its potential exposure, Banc One continually monitored its mark- to-market exposure to each counterparty. Its total exposure to any entity, whether through derivatives or direct lending, was limited by clear policy guide-lines. In addition, to protect itself against the default of a swap counterparty, Banc One required its counterparties to post collateral, in the form of bank- eligible securities or cash, against the bank's expo-sure.[22] Invest-ment bank counterparties posted col-lateral at the initiation of the swap equal to Banc One's possible losses from an extreme one-month move in interest rates.[23] All counterparties were required to post additional collateral as the market value of the swap changed over time.[24] This practice meant that Banc One was not exposed to swap payments for which it did not have collateral, and were the counterpart)' to default, the mark-to-market collateral would allow the bank to enter into a new swap that was economically identical to the one that had defaulted. Banc One's counterparties-and its exposures to each-are shown in Exhibit 6. Banc One's collateral requirements were unique, as most large money-center banks and commercial banks were extremely reluctant to post any kind of collat-eral for swaps, regardless of the counterparty. Yet, because of the magnitude of its derivatives portfolio and because of its solid credit rating, Banc One was almost always able to secure such collateral agree-ments, even from AAA-rated counterparties.

Controlling the Asset and liability Management Process

Banc One's careful handling of counterparty risk was indicative of its long-standing, well-defined investment policies. In late 1993, the investment policies of many banks (including Banc One), and especially their use of derivatives portfolios, came under public scrutiny.

In mid-1993, a consortium of leading financial service firms, known as the Group of Thirty, released a report in which it recommended a set of practices that all derivatives dealers and users follow to ensure that these instruments were used prudently. This report was commonly seen as a proactive effort at self-regulation to fend off governmental regulation of derivatives. Later that year, in October, the U.S. Comptroller of the Currency, the regulator of na-tional banks, issued its own set of guidelines for the use of swaps. The guidelines focused on the role of senior management and boards of directors in ensuring that users of swaps acted safely. The report charged banks with managing market risk, counterparty credit risk, liquidity risk, and opera-tions and systems risk while remaining mindful of the impact of swaps on the banks' capital base and accounting. Politicians seized on the issue and made their own statements concerning the swap market. The statements of the industry, regulators, and politicians pushed the banking sector's use of deriva-tives onto the front pages of leading newspapers and made the issue, one of general interest.

This newfound .interest in the management of derivatives positions came as no surprise to Banc One. For years, senior management had made the prudent use of derivatives and ocher investments, as well as manage-ment of its assets and liabilities, a top priority. Its Asset and Liability Management Commit-tees (ALCOs) were responsible for establishing and implementing policies relating to asset and liability management. The process was governed by a 70-page policy document, updated in April 1993, which outlined an exact system of control and oversight of the bank's asset and liability manage-ment policies, including its management of swaps, an integral part of its investment portfolio. The ALCO process was a system for consistently managing interest rate risk, credit risk, funding risk, and capital adequacy. A committee of the most senior bank executives re-viewed and ratified major investment decisions, recommended changes to existing policy, and moni-tored compliance with policy guidelines.

The ALCO process consisted of regular meet-ings at several levels of the bank. Affiliate banks reviewed their cash position and funds management activities daily. For each state, asset and liability committees were established to monitor that state's activities. At the corporate level, three committees met weekly or monthly to monitor and oversee the overall asset and liability system: the corporate funds management activity committee; the working ALCO committee, which included Lodge, McCoy, and many other senior executives; and the corporate ALCO committee, which included the working ALCO as well as the chairmen of Banc One's holding companies and its chief credit officer. The operation of the MICS system made timely and appropriate information available to each committee.

All policy decisions regarding Banc One's earn-ings sensitivity were made at the corporate level. Furthermore, the firm's investment activities, in-cluding both securities and swaps, were executed at the corporate level by CIO Dick Lodge and his group. Thus, the affiliate and state ALCO groups monitored local deposit and lending activities and their impact on the units' liquidity and interest rate exposure. Corporate ALCO activities overlaid in-vestments and derivatives onto the aggregated ac-tivities of the local banks in order to manage the bank's overall exposure.

When it was established in 1986, the bank's policy was to stay within a 5% earnings sensitivity boundary for an *immediate* 1% shock to interest rates. However, Lodge had recently persuaded the working ALCO committee that such a shock was unrealistic. He believed the committee should in-stead focus on the impact of *agradual* 1% in the level of interest rates during the year (i.e. rates would slowly rise 1%, so that on average they would have risen *112%).* The working ALCO committee agreed to this

Earnings Sensitivity	Policy	Nov. 1993 Banc One Position
1st-year impact for a +1% rate change	(4.00)%	(3.30)%
1st-year impact for a +2% rate change	(9.00)%	(8.00)%
1st-year impact for a +3% rate change	(1;.00)%	(13.20)%
2nd-year impact for a + 1 % rate change	(4.00)%	(1.30)%
2nd-year impact for a + 2% rate change	(9.00)%	(7.90)%
.		
1st-year impact for a -1% rate change	(4.00)%	4.00%

change, and it also set a new boundary for the bank at 4% sensitivity. In addition, the committee set other guidelines:

Within these strategic guidelines, Lodge was permitted, with the working ALCO group's approval, to make tactical decisions on exactly what the bank's earnings sensitivity should be. Although there were several guidelines and Lodge had to comply with each one, both he and the ALCO groups focused mainly on the first-year impact of a gradual 1% change in rates because they believed it was unlikely that interest rates would change much more than 1% in the coming year.

a. Average yield received on investment portfolio (excluding swaps J. For projected period assumes no new investments made.

b. Average scale received on receive-fixed swap portfolio. For projected period assumes no new positions added.

Source: Bane One Corporation.

In November 1993, if it did not have its $12 billion in fixed-rate investments and $22 billion in receive-fixed swaps, the bank' would have been 13% asset sensitive. With them, it was positioned to be 3.3% liability sensitive. This conscious decision to be modestly liability sensitive was the bank's strategic exposure to interest rates. *As* Lodge ex-plained, "Banks are paid to be liability sensitive," meaning that t?e yield curve was almost always upward-sloping. By having a controlled amount of .

long-term, fixed-rate, income-producing assets ex-ceeding its short-term, floating-rate liabilities, the bank could earn the interest differential as long as the yield curve remained upward-sloping and did not shift up dramatically. However, this net position left the bank liability sensitive as a rise in rates would reduce its income.

Although a sudden rise in rates would depress the bank's earnings, the investment portfolio was set up so that this exposure was controlled. Specifically, the swaps in place were level over the next year, but would virtually all mature within two years. Thus, if the bank did not add new swaps to its position, its existing swaps would fall to $17.5 billion by year-end 1994 and $3.6 billion by year end 1995. Its projected earnings sensitivity would drop to -.2% by the end' of 1994, effectively making its earnings unaffected by interest rate swings, and the bank would be asset sensitive by 1995. See Exhibit 7.

Although the bank focused primarily on the impact of interest rates on its earnings, the ALCa committee also examined the effect of interest rates on the value of the firm and its common equity. The asset and liability database allowed it to measure the duration of assets and liabilities. Lodge's figures for the bank's key duration measures,[25] as of September 30, 1993, were 1.73 years for on- and off-balance sheet assets and 1.51 years for its liabilities. Because the difference between assets and liabilities was a residual equity account, Lodge could also calculate a rough duration of equity (by weighting each category by its total dollar amount). *As* of September 1993, residual equity had a duration of +4.00 years. For each 1% rise in rates, this duration measure suggested that Banc One's equity value would drop by 4.0%. As interest rates rose, its slightly longer duration asset base would decline in value faster than its shorter duration liabilities, leading to a magnified drop in the market value of its equity.

As of September 1993, Banc One had $37.7 billion in notional volume of interest rate, swaps on its books. Both Lodge and McCoy felt that the bank had drawn some of its unwanted attention because its swap portfolio had grown so dramatically. One analyst identified Banc One as having the second- largest growth in an existing swap portfolio of all regional banks. At the end of 1990, Banc One had only $4.7 billion in swaps on its books. This figure had grown to 513.5 billion at the end of 1991 and $21.0 billion at the end of 1992. Looking forward, Banc One saw continued growth in its swap portfo-lio as long as its earnings grew, it continued to acquire banks that were more asset sensitive than itself, and it faced an upward-sloping yield curve.

Disclosure

As of November 1993, the Financial Accounting Standards Board (FASB) required minimal disclo-sure of the details of a company's swap portfolio because swaps were classified as off-balance-sheet items. Generally, the total notional volume of swaps was reported as a footnote to reported financial state-ments. Under accounting guidelines, though, notional volume had EO include *all* swaps , regardless of their purpose or whether they offset one another. Thus, if Banc One entered into a $100 million receive-fixed swap and then a $100 million basis swap to adjust the floating-rate index it paid, the swaps would be reported as $200 million of notional amount, even though they economically replicated only $100 million of a fixed-rate investment. Like-wise, if it entered into a $100 million pay-fixed swap and then entered into an exactly offsetting receive- fixed swap, it would report $200 million in swaps.

Even though FASB required minimal swap disclosure, Banc One had voluntarily disclosed additional information, consis-tent with its reporting policies. In addition to reporting the total notional volume of swaps on its books, it reported the unrealized net gain or loss on its swap portfolio. Banc One's disclosures of its swaps activities for 1993 are shown in Appendix 1.

The Meeting

As Banc One's earnings grew, so too did its swap position. With its growing swap portfolio, it caught the attention of bank analysts. Some ap-plauded the bank's use of swaps to manage its interest rate exposure. Other-more vocal-analysts were critical, accusing Banc One of using swaps to inflate earnings, overstate capital ratios, and offset declines elsewhere in the bank. These critics saw the rapidly growing swap positions as heading out of control. One analyst was quoted as saying of the bank's swap activities, "Does that look like hedge activity? They use this stuff to keep the game going." A few analysts had downgraded the stock.

Though it was impossible to pin the recent decline in Banc One's stock price solely on its growing derivatives portfolio, both insiders and outsiders felt that the $10 drop in its stock price was due in large part to the market's reaction to the bank's use of derivatives. One analyst supportive of the company wrote:

One likely reason for the price weakness is a recent focuses on Bane One's liberal use of derivatives to achieve its asset/liability management goals. Since derivatives are relatively new financial instruments, and since their use requires a high degree of financial sophistication and quantitative expertise, there is an understandable aversion to them on the part of many investors.. .Although (Bane One's swap po-sition) is a large notional amount for a regional bank, we think Bane One's use of derivatives has been prudent.[26]

As the meeting between McCoy and Lodge began, McCoy voiced his concern about Banc One's falling stock price. From his perspective, he and Lodge faced a dilemma. On the one hand, he felt that swaps were hurting shareholder value because the investment community did not understand how they were being used. On the other hand, he believed that they were an invaluable tool in managing risk. Given the distance between his beliefs and. what he was hearing from the market, he wondered what, if anything, the bank should do.

In an attempt to answer this question, McCoy and Lodge discussed three possible options. First, they could do nothing and hope that Banc One's stock price would recover over time as investors realized that derivatives were actually helping the bank manage interest rate and basis risk. Second, they could abandon or severely limit their deriva-tives portfolio. Third, they could attempt to educate investors about how they used derivatives. Their most recent quarterly disclosure gave the market a great deal of data on the bank's swap portfolio, but perhaps even more information might dispel the misconceptions. What information would the mar-ket want to see? And how could Banc One credibly present it so as to convince its skeptics and educate swap novices? Perhaps analysts would understand Banc One's ALCO process and use of swaps if they could compare the bank to a hypothetical Banc One that had no swaps or no investments. In preparation for the meeting, Lodge and his staff prepared a set of analyses showing this comparison (see Appendix 2).

None of the alternatives was riskless. Doing nothing might give the impression that the bank was hiding something, thereby confirming investors' worst suspicions. If it caused Banc One's stock price to stay low or fall even further, the bank's ability to continue its stock acquisitions would be jeopar-dized. Eliminating its derivatives portfolio would leave the bank with ,greater interest-rate exposure and few tools to manage it. Disclosing even more information was not a guaranteed solution. In draw-ing even greater attention to its derivatives portfolio,

the bank might raise investors' concerns or increase .their confusion.

Appendix 1: Banc Ones 1993 Disclosure Of Its Interest Rate Management Activities (10-q-fillings)

Panel A: 1993 First Quarter

BANC ONE manages interest rate sensitivity within a very small tolerance through the use of off-balance sheet interest rate swaps and other instruments, thereby minimizing the effect of interest rate fluctuations on earnings and market: values. The use of swaps resulted in BANC ONE being slightly liability-sensitive at March 31, 1993, countering the natural

Panel B: 1993 Second Quarter

BANC ONE manages interest rate sensitivity within a very small tolerance through the use of off-balance sheet interest rate swaps and other instruments, thereby minimizing the effect of interest rate fluctuations on earnings and market values. The use of swaps resulted in BANC ONE being slightly liability-sensitive at June 30, 1993, adjusting the natural ten-dency to be asset-sensitive. Swaps increased interest income by $59 million and $113 million for the three and six month periods ending June 30, 1993 as compared to $46 million and 595 million for the same periods in 1993. Swaps decreased deposit and other borrowing cost by $48 million and 596 million for the three and six month periods ended June 30, tendency to be asset-sensitive. The use of swaps to manage interest rate sensitivity increased interest income by $54 mil-lion and 550 million, and decreased interest expense by $47 and $34 million in the first quarter of 1993 and 1992, respec-tively. The notional amount of swaps increased from $8.3 billion to $23.4 billion from March 31, 1992 to March 31, 1993.

1993, compared to decreases of S45 million and 580 million for the same periods in 1992. The notional amount of swaps increased to $31.5 billion from $20.8 billion and $18.4 billion at December 31, and June 39, 1992, respectively. Accruing fixed rate swaps represented $17.4 billion, $10,5 billion and $11.2 billion for the same respective periods.

Along with the second quarter report, Bane One made available to its investors a 100 page brochure entitled Bane One Corporation Asset and Liability Management. This brochure described how the corporation uses swaps and other deriva-tives to maintain its strong capital position, manage its liquid-ity, and manage the bank's interest rate exposure.

Panel C: 1993 Third Quarter

The following information supplements Management's Discussion and Analysis in Part 1. The notional amount of swaps shown below represents an agreed upon amount on which calculations of interest payments to be exchanged are based. BANC ONE's credit exposure is limited to the net difference between the calculated pay and receive amounts on each transaction which are generally netted and paid quarterly. BANC ONE's policy is to obtain sufficient collateral from swap counterparties to secure receipt of all amounts due, At September 30, 1993, the market value of interest rate swaps and the related collateral was approximately $5365 million and $623 million respectively. As indicated below the notional value of the interest rate swap portfolio increased from 521 billion to 538 billion during the nine months ended September 30, 1993, This increase was primarily associated with swaps acquired to replaced fixed rate, on-and off-balance sheet instruments which have or will mature or amortize and to manage interest rate risk in newly acquired affiliates. These new affiliates did not use swaps to manage their exposure to interest rate risk to as great a degree as BANC ONE. Exposure to interest rate risk is determined by simulating the impact of prospective changes in interest rates on the results of operations. Management seeks to insure that over a one-year horizon, net income will not be impacted by more than 4 percent and 9 percent by a gradual change in market interest rates of 1 percent and 2 percent, respectively. At

December 31, 1992, a 2.3 percent reduction in forecasted earnings would have resulted from a gradual 1 percent increase in market rates. Due to the increase in the notional value of the swap portfolio noted above, the sensitiv-ity to such a rate increase changed to 3.8 percent at Septem-ber 3D, 1993. BANC ONE believes that both on-balance sheet securities and off-balance sheet derivatives may be used inter-changeably to manage interest rate risk to an acceptable level. Various factors are considered in deciding the appropriate mix of such securities and derivatives including liquidity, capital requirements and yield,

During the nine months ended September 30, 1993, BANC ONE entered into 53.8 billion notional amount of basis swap contracts where payments based on the prime rate and LIBOR are exchanged. The variable rate used in the non-basis swap contracts entered into by BANC ONE are based on LIBOR, while many of the variable rate assets being synthetically altered are based on the prime rate. Basis swap contracts, therefore, improve the degree to which the swap portfolio acts as a hedge against the impact of changes in rates on BANC ONE's results of operations.

The table below summarizes by notional amounts the activ-ity for each major category of swaps. For all periods presented, BANC ONE had no deferred gains or losses relating to termi-nated swap contracts. The terminations shown in the follow-ing table for the year ended December 31, 1991 resulted in losses of 51.8 million which were recognized during that year in accordance with BANC ONE's accounting policy at that time. The terminations in 1993 related to swaps which had been carried at market value and, therefore, resulted in no deferred gain or *loss* at termination.

$ (millions)	Received Fixed	Pay Fixed	Basis	Forward Starting	Total
Balance, December 31, 1990	**$ 3,114**	**$ 937**	**$550**	**$117**	**$4,719**
Additions	9,797	509			10,306
Maturities / Amortizations	(1,171)	(322)			(1,493)
Terminations	(3,102)				(3,102)
Forward Starting-Becoming Effective	117			117	
Acquisition and other (net)	2,764	277			3,041
Balance December 31, 1991	**11,519**	**1,401**	**550**		**13,470**
Additions	2,002	501		11,656	14,159
Maturities/Amortizations	(6,059)	(182)	(350)		
Terminations					
Forward Starting-Becoming Effective	3,201	1,005		(4,206)	
Acquisitions and Other (net)	289	(296)			(7)
Balance, December 31, 1992	**10,952**	**2,429**	**200**	**7,450**	**21,031**
Additions	4,428	1,237	3,800	12,000	21,465
Maturities/Amortizations	(3,545)	(861)	(204)		(4,610)
Terminations	(250)	(250)			(500)
Forward Starting – Becoming Effective	10,480			(10,480)	
Acquisition and Other (net)	450	15	20	(150)	335
Balance, September 30, 1993	**$22,515**	**$2,570**	**$3,816**	**$8,820**	**$37,721**

The table below summarized expected maturities and weighted average interest rates to be received and paid on the swap portfolio at September 30, 1993: A key assumption in the preparation of the table is that rates remain constant at Septem-ber 30, 1993 levels. To the extent that rates change, both the periodic maturities and the variable interest rates to be received or paid will change. Such changes could be substantial. The maturities change when interest rates change because the swap portfolio includes $23.6 billion of amortizing swaps. Amortiza-tion is generally based on certain interest rate indices.

EXPECTED MA TURITY								
	4th Quart.							
$ (millions)	1993	1994	1995	1996	1997	1998	All Other	Total
Receive Fixed Swaps								
Notional Amount	$2,436	$9,096	$8,880	$1,050	$90	$46	$917	$22,515
Weighted Average								
Receive Rate	7.58%	6.00%	5.34%	6.02%	7.24%	6,22%	6.81%	5.95%
Pay Rate ………………………	6.64	3.28%	3.23	3.36	3.24	3.19	3.54	3.19
Pay Fixed Swaps								
Notional Amount ………………	$627	$970	$318	$272	$267	$109	$7.	$2,570
Weighted Average								
Receive Rate	3.25%	3.39%	3.33%	3.26%	3.44%	3.41%	3.31%	3.34%
Pay Rate	6.64	5.86	5.00	5.76	6.07	5.30	8.82	5.96
Basis Swaps								
Notional Amount …..................	0	0	0	$2,200	$1,600	$16	0	$3,816
Weighted Average								
Receive Rate	0.00	0.00	0.00	3.22%	3.27%	3.20%	0.00	3.24%
Pay. Rate	0.00	0.00	0.00	3.33	3.34	4.80	0.00	3.34
Forward-Starting"								
Notional Amount	$500	$100	$6,720	$1,500	0	0	0	$8,820
Weighted Average								
Receive Rate ……………………	7.20%	5.74%	4.98%	5.68%	0.00	0.00	0.00	5.24%
Pay Rate	3.38	3.38	3.38	3.38	0.00	0.00	0.00	3.38
Total..	$3.563	$10,166	$15.918	$5.022	$1,957	$171	$924	$37.721
	6.77%	5.75%	5.15%	4.54%	3.47%	4.14%	6.78%	5.33%
	3.88	3.53	3.33	3.48	3.71	4.69	3.58	3.49

In preparation for his meeting with McCoy, Dick Lodge asked his staff to prepare a simplified set of Banc One financials that could communicate the essence of the bank's financial state-ments and the underlying economics of their business. This stylized set of financials would show the basic earnings sensi-tivity faced by the bank, and how it used swaps to solve this problem. The simplified model would also demonstrate the impact of the bank's derivative activities on its accounting ratios, such as its net interest margin as

Rate		Banc One Stylized	Twin A (No investment Activities)	Twin B (Swap on Balance Sheet)		Rate	Banc One (Stylized)	Twin A (No investment Activities	Twin B (Swap on Balance Sheet)
BALANCE SHEET ($ IN BILLIONS)					• Additional Treasury securities	4.30%	0.00	0.79	0.00
Assets					Total interest income		5.47	6.26	5.66
Floating-rate assets					Interest expense from:				
• Variable-rate loans		533.8	533.8	533.8	• Retail deposits	3.27%	0.63	0.63	0.63
• Additional money					• Wholesale deposits	3.09%	0.27	0.27	0.27
market assets		0	0	31.8	• Additional wholesale deposits	3.09%	0.00	0.57	0.57
Fixed-rate Assets					• Fixed core deposits	3.57%	0.85	0.85	0.85
• Fixed-rate loans		18.6	18.6	18.6	• Large deposits	3.57%	0.08	0.08	0.08
• Fixed-rate investments		13.4	13.4	0	Total interest expense		1.83	2.40	2.40
• Additional Treasury securities		0	18.4	0	Income from Swaps (6)	2.50%	0.46	0.00	0.00
Other assets		8.4	8.4	8.4	Net interest		4.09	3.85	3.25
Total Assets		$74.2	$92.6	$92.6	Non-interest expense		2.37	2.37	2.37
NOTE: Earning Assets (1)		65.8	84.2	84.2	Taxable earnings		1.72	1.48	0.88
					Taxes	34.00%	0.59	0.50	0.30
Liabilities and Equity					Net income		1.14	0.98	0.58
Floating-rate liabilities									
• Retail deposits		19.3	19.3	19.3	PERFORMANCE MEASURES				
• Wholesale deposits (2)		8.8	8.8	8.8	Net interest margin (7)		6.22%	4.58%	3.86%
• Additional wholesale					Net interest margin				
deposits (3)		0.0	18.4	18.4	(excluding swaps) (8)		5.52%	4.58%	3.86%
Fixed-rate liabilities					Return on assets		1.53%	1.06%	0.63%
• Fixed core deposits (4)		23.8	23.8	23.8	Equity/Assets (9)		8.56%	6.86%	6.86%
• Large time deposits		2.3	2.3	2.3	Return on Equity (10)		17.89%	15.42%	9.19%
Other liabilities		13.4	13.4	13.4	Dependence on large				
Total liabilities		67.6	86.0	86.0	liabilities (11)		15.0%	33.5%	-5.4%
Preferred shares		0.3	0.3	0.3	Risk-adjusted assets 02)		$63.2	$63.1	$74.7
Common shares		6.4	6.4	6.4	Tier I capital/risk -adjusted				
Total		$74.2	$92.6	$92.6	assets (13)		10.4%	10.5%	8.8%
					Earnings sensitivity (14)		-3.30%	-3.30%	12.88%
OFF-BALANCE-SHEET ITEMS									
Swaps (5)		$18.4	$0.0	$0.0	SUMMARY				
• INCOME STATEMENT					Earnings		High	Better	Low
Interest Income from:					Capital		High	Low	Low
• Variable-rate loans	7.32%	$2.47	$2.47	$2.47	Risk Capital		Good	High	Low
• Additional money					Liquidity		Good	Low	High
market assets	3.50%	0.00	0.00	1.11	Earnings Sensitivity		Liability	Liability	Very
• Fixed-rate loans	11.13%	2.07	2.07	2.07			Sensitive	Sensitive	Asset
• Fixed-rate investments	6.88%	0.92	0.92	0.00					Sensitive

well as its returns on assets and equity. Moreover, the simplified books would show how swaps affected the bank's dependence on large shor-t term borrowings as well as demonstrate how the bank's swap portfolio affected the amount of risk-adjusted capital it held.

In order to explain the role that swaps played at Banc One, Lodge and his staff felt it might be instructive to compare Banc One with two hypothetical twin banks whose investment policies differed from its own. The first twin was like Banc One in all regards but one. This hypothetical bank brought its swaps onto the balance sheet by replacing the notional princi-pal of the' its receive-fixed swaps with investments in fixed- rate securities27 funded by variable-rate borrowings. Because Banc One's receive-fixed swaps were similar to an investment in fixed-rate securities funded by floating-rate borrowings, t?is twin would have similar interest rate exposure to Banc One. However, it would differ in its accounting performance, de-pendence .on large liabilities, and capital levels.

A second twin would follow yet another investment strategy. In place of Banc One's fixed-rate investments, this twin would invest in floating-rate loans and investments. In place of Banc One's swaps, it would invest in floating-rate assets financed by floating-rate deposits. The second twin more closely resembles a bank that did not manage its interest rate sensitivity. .

The hope was that these simple projections would demon-strate to investors how the bank's investment activities, but especially its derivatives activities, affected its earnings sensi-tivity, accounting results, liquidity, and capital needs.

1. Earning assets include loans and investments.
2. "Wholesale" deposits represent liabilities to other financial institutions, eg., federal funds borrowings.
3. For both twin banks, additional needs for fund would be met by borrowing from other
   financial institutions.
4. Fixed core deposits are the "Sticky fixed- deposits. Their rates may change with market rates (at bank management', discretion), but they are relatively stable in volume as rates change.
5. Represents only the swaps in which Banc One receives and the current floating rate. Does not include Banc One's basis swap.
6. Represents the difference between the fixed rate that Banc One receives and the current floating rate. Does not include Banc One's basis swaps.
7. Net interest (including income from swaps) divided by earning assets.
8. Net interest (excluding income from swaps) divided by earning assets.
9. Common equity / assets.
10. Return to common equity.
11. Equals (large time deposits + wholesale deposits – money market assets) / (earning assets – money market assets). Represents an estimate of the liabilities that the bank might be called on to honour immediately, net of its assets that could be liquidated immediately.
12. Calculated by applying the BIS capital weights to each assets category.
13. Banc one's equity divided by its risk adjusted assets.
14. Represents the percentage change in the coming year's net income in response to a gradual 1% rise in interest rates over the coming year. In this model, a gradual 1% rise in rates is the same as an immediate. 5% increase in rates. The earnings sensitivity for a 1% rise. This is because of the amortization schedule of the bank's swap contract as well as the nature of the other bank assets and liabilities. Furthermore, a 1% fall in rates would not necessarily produce the same earnings sensitivity. Bank One estimated that a 1% drop in rates would lead to a 4.0% increased in earnings as compared to a 3.3% decline in earnings for a 1% rate increase.
15. Hoopoe Corp. wants to borrow 100 million U.S. dollars at a fixed rate with a maturity of 5 years. It calculates that it can make a eurobond issue with the following terms:

Interest	$10^{5/8}$% payable annually
Maturity	5 years
Commission	1 $^{1/8}$ %
Agency Fee	15% on coupon
	.075% on principal
Issue expenses:	.2%

A bank has presented Hoopoe with a proposal for a Swiss franc issue combined with a currency swap in U.S. dollars. The proposed terms for the Swiss franc issue are:

Amount	200 million Swiss francs
Interest	5 $^{3/8}$% annually
Maturity	5 years
Commission	2.8%
Agency Fee	.75% on coupon
	.30% on principal
	.2%

The counterparty of the swap would raise fixed dollars on the following terms:

Amount:	100 million U.S. dollars (equivalent to 200 million Swiss francs)
Interest:	10 $^{5/8}$ % annually
Maturity:	5 years
Commissions:	1.8%
Agency expenses:	.15% on coupon
	.075% on principal
Issue expenses:	.2%

The counterparty would be happy with an all-in cost in Swiss francs of 6.4 percent.

a. Which alternative should Hoopoe undertake? (Ignore credit risk in your analysis.)

b. Suppose that you are the corporate finance manager of Hoopoe. Discuss the credit risk issues involved in the alternatives.

UNIT-VI

CASH FLOW ENGINEERING & FORWARD CONTRACTS

Objectives:

- On completion of this lesson you will be able to understand of how cash flow can be replicated and then repackaged to create synthetic instruments.

Dear Friends, all financial instruments can be visualized as bundles of cash flows. They are designed so that market participants can trade cash flows that have different characteristics and different risks. This chapter uses forwards and futures, to discuss how cash flows can be replicated and then repackaged to create synthetic instruments.

It is easiest to determine replication strategies for linear instruments. We show that this can be further developed into an analytical methodology to create synthetic equivalents of complicated instruments as well. Thus we arc less concerned with specific types of synthetics than with, methods for their construction. This analytical method will be summarized by a (contractual) equation. After plugging in the right instruments, the equation with yield the synthetic for the cash flow of interest. Throughout this chapter, we assume that there is no default risk and we discuss only static replication methods. Positions are taken and kept unchanged until expiration, and require no rebalancing.

What is a Synthetic?

The notion of a synthetic instrument, or replicating portfolio, is central to financial engineering. We would like to understand how to price and hedge an instrument, and learn the risks associated with it. To do this we consider the cash flows generated by an instrument during the lifetime of its contract. Then, using other simpler, liquid instruments, we form a portfolio that replicates these: cash flows exactly. This is called a replicating portfolio and will be a synthetic of the original instrument. The constituents of the replicating portfolio will be easier to price understand and analyze than the original instrument.

In this lesson, we start with synthetics that can be discussed using forwards and futures and money market products. At the end we obtain a contractual equation that can be algebraically manipulated to obtain solutions to practical financial engineering problems.

Cash Flows

We begin our discussion by defining a simple tool that plays an important role in the first part of this book. This tool is the graphical representation of a cash flow.

By a cash flow, we mean a payment or receipt of cash at a specific time, in a specific cur-rency, with a certain credit risk. For example, consider the cash flows in Figure 3-1. Such figures are used repeatedly in later lessons, so we will discuss them in detail.

Example

In Figure 3-1a we show the cash flows generated by a loan. Multiplying these cashflows by -1 converts them to cash flows of a deposit, or depo. In the figure, the horizontal axis represents time. There are two time periods of interest to us, denoted by symbols t_0 and t_1. The t_0 represents the time of a $100 cash inflow. It is represented by a rectangle, above the line. At time t_1, there is a cash outflow, since the rectangle is placed below the line and thus indicates a debit. Also note that the two cash flows have different sizes.

We can interpret Figure 3-1 a as cash flows that result when a market participant borrows 100 USD at time to and then pays this amount back with interest as 105 USD, where the interest rate applicable to period $[t_0, t_1]$ is 5%.

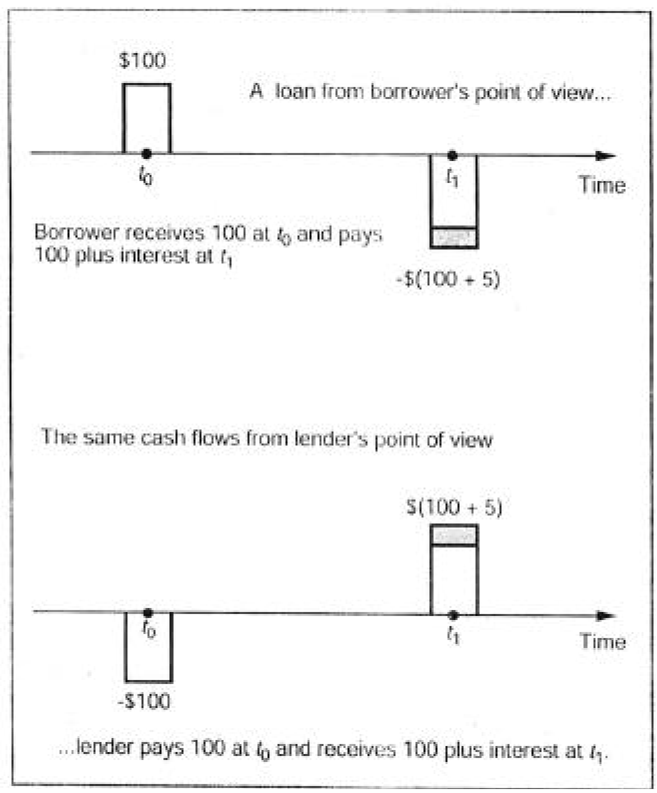


Figure 1a

It is important to realize that the top portion of Figure 1a shows the cash flows from the borrower's point of view. In fact, every financial transaction has at least two coun-terparties. Thus, if we look at the same instrument from the lender's point of view, we will see an inverted image of these cash flows. The lender lends $100 at time to and then receives the principal and interest at time t_1. The bid-ask spread suggests that the interest is the asking rate.

Finally, note that the cash flows shown in Figure 1 a do not admit any uncertainty. Since, both at time t_0 and time- t_1 cash flows are represented by a single rectangle with known value. If there were uncertainty about either one, we would need to take, this into account in the graph.

For example, if there was a default possibility on the loan repayment, then the cash flows would be represented as in Figure 1 b. If the borrower defaulted, there would be no payment at all. At time t_1, there ore two possibilities. The lender either receives $105 or receives nothing.

Cash flows have special characteristics that can be viewed as *attributes.* At all points in time, there are market participants and businesses with different needs in terms of these attributes. They will exchange cash flows in order to reach desired objectives. This is done by trading financial contracts associated with different cash flow attributes. We now list the major types of cash flows with well-known attributes.

Cash Flows in Different Currencies

The first set of instruments devised in the markets trade cash flows that are identical in every respect except for the currency they are expressed in.

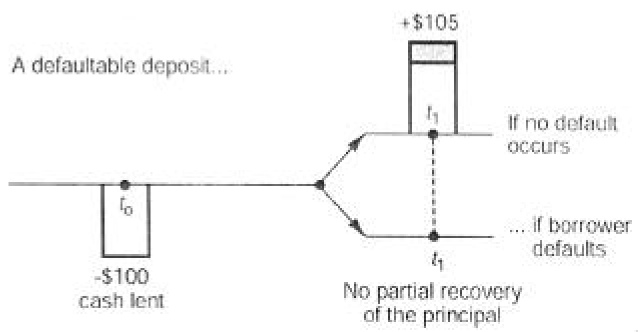


Figure 1b

In Figure 2, a decision maker pays l00 USD at time to and receives l00 e_{to} units of Euro at the same time. This a spot FX deal, since the transaction takes place at time to. The e_{to} is the spot exchange rate. It is the number of Euros paid for one USD.

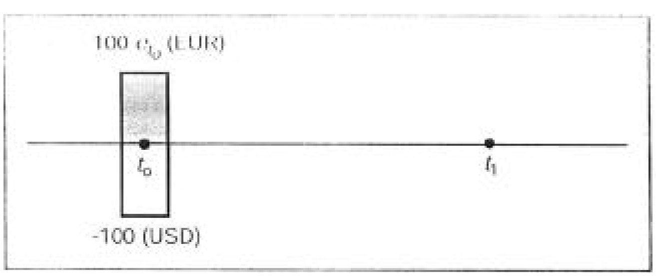


Figure 2

Cash Flows with Timing Differences

One may want to exchange a cash How that belongs to date t_0 against a cash flow that belongs to a different time period. h. An example is shown in Figure 3-3. The market participant makes a cash payment at time to and receives a cash flow in the same currency at time t_1. The difference between the sizes of the two cash flows is the interest earned during this period. Every loan or deposit will fall into this category.

Note that these loans can be in terms of gold, silver, wheat, or other commodities, as well as in specific currencies.

Cash Flows with Different Market Risks

If cash flows with different market risk characteristics are exchanged we obtain more complicated instruments than a spot FX transaction or deposit. Figure 3-4 shows an exchange of cash flows that depend on different market risks. The market practitioner makes a payment proportional to L_{t1} percent of a notional amount N against a receipt of F_{t0} percent of the same N. The subscripts indicate that. L_{t1} is an unknown, floating rate at time to that will be learned at time t_1. The Ft_0 on the other hand is set at time t_0 and is forward interest rate. The cash flows are exchanged at time t_2 and involve two different types of risk. Instruments that are used to exchange such risks are often referred to as swaps. They exchange a floating risk against a fixed risk. Swaps are not limited to interest rates. For example a market participant may be willing to pay a floating (i.e. to be determined) oil price and receive a fixed oil price. One can design such swaps for all types of commodities.

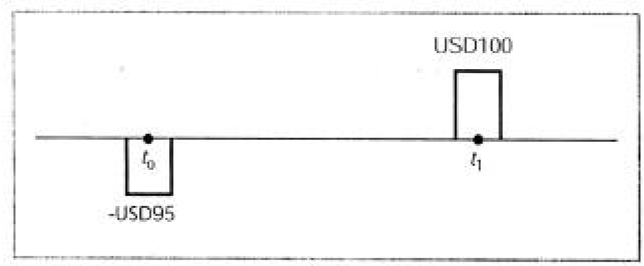


Figure 3

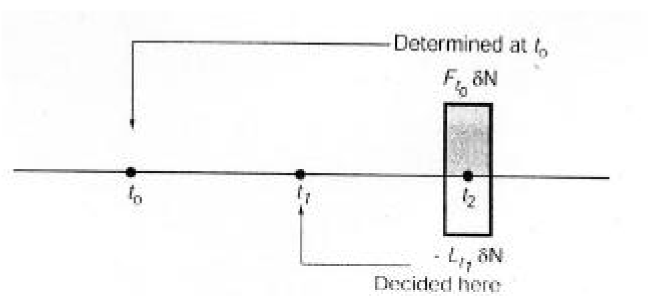


Figure 4

Cash Flows with Different Credit Risks

The probability of default is different for each borrower. Exchanging cash flows with different credit risk characteristics leads to credit instruments.

In Figure 3-5, a counterparty makes a payment that is contingent on the default of a decision maker against the guaranteed receipt of a fee. Market participants may buy and sell such cash flows with different credit risk characteristics and thereby adjust their credit exposure. For example, AA-rated cash flows can be traded against BBB-rated cash flows.

Cash Flows with Different Volatilities

Instruments that exchange cash flows with different volatility characteristics are rather new. Figure 3-6 shows the case of exchanging a fixed volatility at time t_2 against a realized (floating) volatility observed during the period, $[t_1, t_2]$. Such instruments are called volatility or Vol-swaps.

Cash Flows with Different Sensitivities

Cash flows arc dependent not only Oil risk factors, but may depend on these risks with different sensitivities. One may want to exchange cash flows with different sensitivities to the same risk factor.

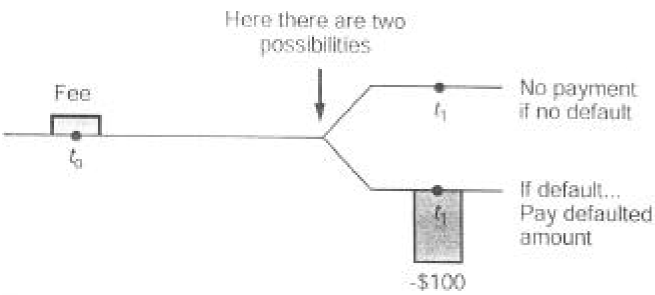


Figure 5

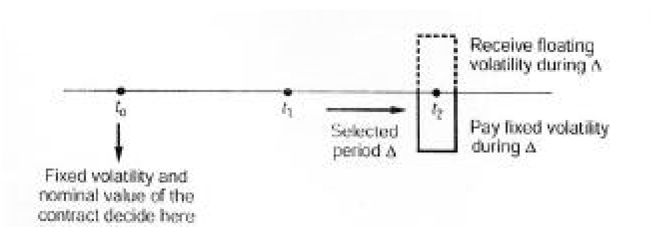


Figure 6

Figure 3-7 shows one example. The price of a 3D-year discount bond has a pronounced curvature, implying that the sensitivity of the bond price to changes in yield is *not* constant. In fact, the 3D-year bond price is a *nonlinear* function of the yield. The same cannot be said for a 2-yearbond. Its price appears to be a (quasi) *linear* function. Instruments can be devised such that *convexity* differences are traded.

Options and bonds are two of the most common instruments that can be used to trade convexity.

Forward Contracts

This chapter deals only with the most elementary cash flow exchanges. We consider *forwards, futures contracts,* and the underlying inter-bank *money markets.* These are the simplest and some of the most liquid instruments. They are ideal for creating synthetic instruments for many rea-sons. Forwards and futures are, in general, linear. They are often very liquid and, in case of currency forwards, have homogenous underlying. Many technical complications are automati-cally eliminated by the homogeneity of a currency. Forwards and futures on interest rates present more difficulties, but a discussion of these will be postponed until the next chapter.

A forward or a futures contract can fix the future selling or buying price of an underlying item. This can be useful for hedging, arbitraging, and pricing purposes. They are essential in creating synthetics. Consider the following interpretation.

Instruments are denominated in different currencies. A market practitioner who needs to perform a required transaction in US dollars normally uses instruments denoted in US dollars. In the case of the dollar this works out fine since there exists a broad range of liquid mar-kets. Market professionals can offer all types of services to their customers using these. On the other hand, there is a relatively small number of, say, *liquid* Swiss Franc (CHF) denoted instruments. Would the Swiss market profes-sionals be deprived of providing the same services to their clients? It turns out that liquid Foreign Exchange(FX) forward contracts in USD/CHF can, in principle, make USD-denomi-nated instruments available to CHF-based clients as well. Instead of performing an operation in CHF, one can first buy and sell USD at t_0, and then use an USD-denominated instru-ment to perform any required operation. Liquid FX-Forwards permit *future* USD cash flows to be reconverted into CHF as of time t_0. Thus, entry into and exit from a different currency is fixed at the initiation of a contract. As long as liquid forward contracts exist, market professionals can use USD-denominated instruments in order to perform operations in any other currency.

As an illustration, we provide the following example where a synthetic zero coupon bond is created using FX-forwards and the bond markets of another country.

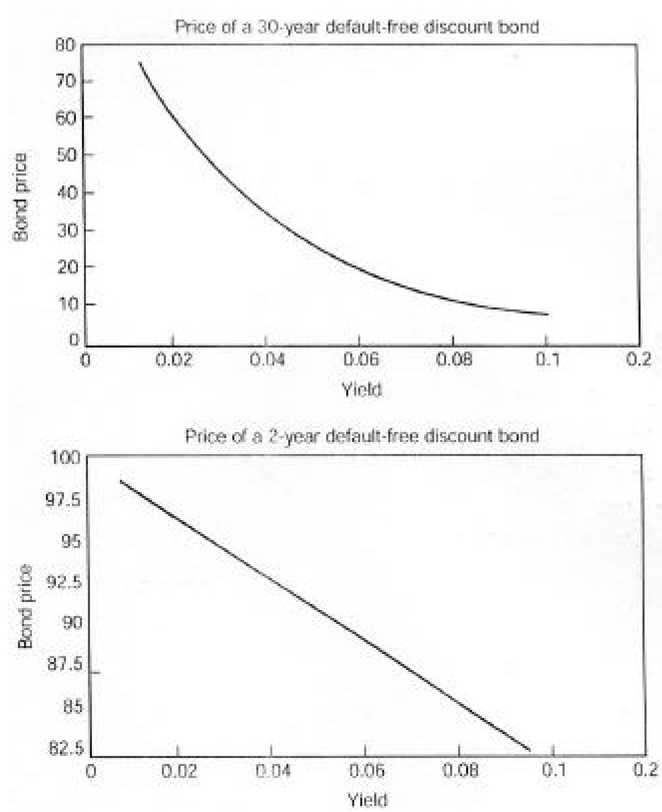


Figure 7

Example

Suppose we want to buy, at time to, a USD-denominated default-free discount bond, with maturity at t_1 and current price B (t_0, t_1). We can do this synthetically using bonds denominated in any other Currency, as long as FX-forwards exist and the relevant credit risks are the same.

First, we buy an appropriate number of, say, Euro-denominated bonds with the same maturity, default risk, and price B(t_0, t_1)^E. This requires buying Euros against dollars in the spot market at an exchange rate $e_{t0.}$ Then, using a forward contract on Euro, we sell forward the Euros that will be received on December 31,2005, when the bond matures. The forward exchange rate is F_{t0}.

The final outcome is that we pay USD now and receive a known amount of USD at maturity. This generates the same cash flows as a USD - denominated bond. This operation is shown in Figure -8.

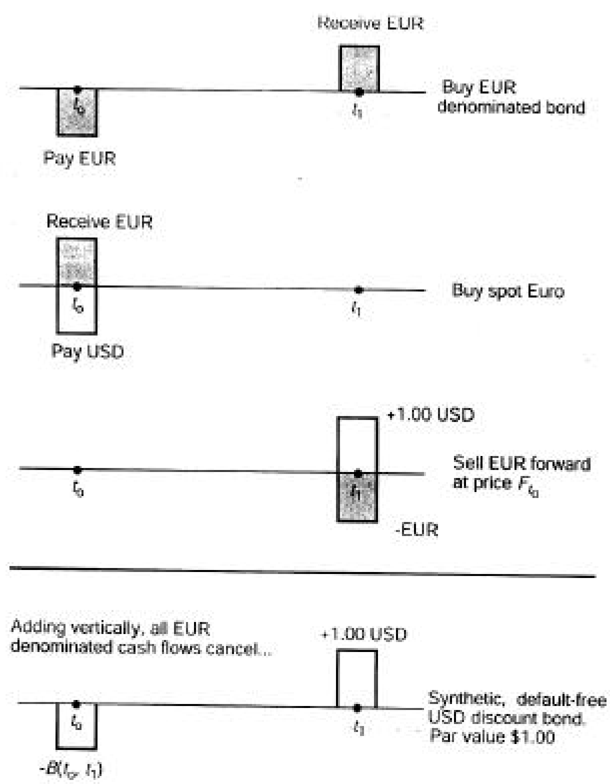


Figure 8

In principle, such steps can be duplicated for any (linear) underlying asset, and the ability to execute forward purchases or sales plays a crucial role here. Before we discuss such operations further, we provide a formal definition of forward contracts.

A forward is a contract written at time to, with a commitment to accept delivery of (deliver) *N* units of the underlying asset at a future date t_1, $t_0 < t_1$ at the *forward price* F_{t0} The current price of the underlying asset S_{t0} is called the spot price and is not written anywhere in the contract, instead, F_{to} is used during the settlement. Note that F_{to} has a t_0 subscript and is fixed at time t_0. At time to, nothing changes hands; all exchanges will take place at time t_l. An example of such a contract was shown in Figure 3-8.

Forward contracts are written between two parties, depending on the needs of the client. They are *flexible* instruments. The size of contract *N*, the expiration date h, and other conditions written in the contract can be adjusted in ways the two parties agree on.

If the same forward purchase or sale is made through an *homogenized contract,* in which the size, expiration date, and other contract specifications are preset if the trading is done in a *formal exchange,* and if there is formal *mark-to-market,* then the instrument is called *a futures* contract.

Positions on forward contracts arc either *long* or *short.* As discussed in Chapter 2, a *long position* is a commitment to accept delivery of the contracted amount at a future date, t_1 at price F_{t0} This is displayed in Figure 3-9. Here F_{t0} is the contracted forward price. As time passes, the corresponding price on newly written contracts will change and at expiration the forward price becomes F_{tl}. The difference. $F_{t1} - F_{t0}$, is the profit or loss for the position taker. Note two points. Because the forward contract does not require any cash payment at ini-tiation, the time-to value is *on* the x-axis. This implies that, at initiation, the market value of the contract is zero. Second, at time t_1 the spot price and the forward price will be the same (or very close).

A *short position* is a commitment to deliver the contracted amount at a future date $t_{1,}$ at the agreed price F_{t0}. The short forward position is displayed in Figure 3-9. The difference F_{to} – F_{t1} is the profit or loss for the party with the short position.

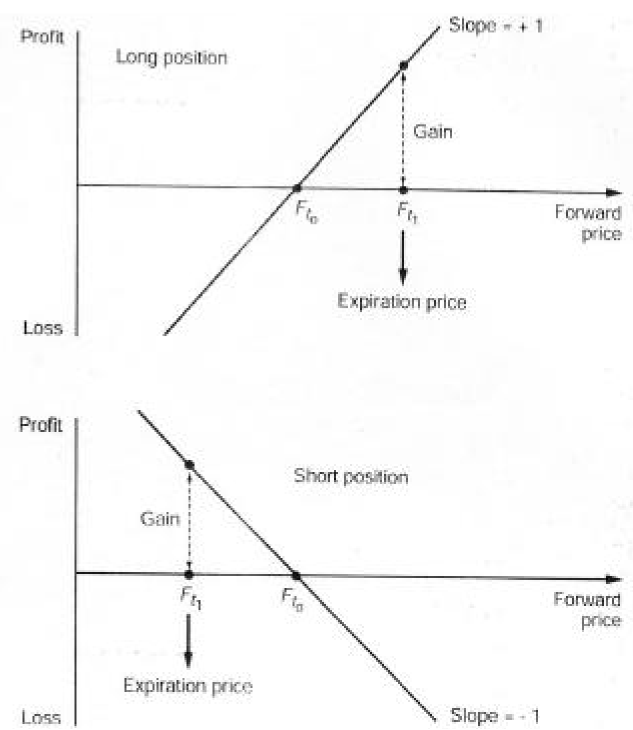


Figure 9

Examples

Elementary forwards and futures contracts exist 011 a broad array of underlyings. Some of the best known are the following:

1. Forwards on currencies. These are called FX-forwards and consist of buying (selling) one currency against another at a future date t_1
2. Futures On loans and deposits. Here, a currency is exchanged against itself, but at a later date. We call these forward loans or deposits. Another term for these is forward-forwards. Futures provide a more convenient way to trade interest rate commitments; hence forward loans are not liquid. Futures on forward loans are among the most liquid.
3. Futures on commodities, e.g., be oil, corn, pork bellies, and gold. There is even a thriving market in futures trading on weather conditions.
4. Futures and forwards on individual stocks and stock indices. Given that one cannot settle a futures contract on an index by delivering the whole basket of stocks, these types of contracts are cash settled. The losers compensate the gainers in cash, instead of exchanging the underlying products
5. Futures contracts 011 swaps. These are relatively recent and they consist of future swap rate commitments. They are also settled in cash. Compared to futures trading, the GTC forward market is much more dominant here.

We begin with the engineering of one of the simplest and most liquid contracts; namely the currency *forwards.* The engineering and uses of forward *interest rate products* are addressed in the next chapter.

Currency Forwards

Currency forwards are very liquid instruments. Although they are elementary, they are used in a broad spectrum of financial engineering problems.

Consider the EUR/USD exchange rate.[2] The cash flows implied by a forward purchase of 100 US dollars against Euros are represented in Figure 3-l0a. At time to, a contract is written for the forward purchase (sale) of 100 US dollars against 100 / F_{to} Euros. The settlement-that is to say the actual exchange of currencies - will take place at time t_1. The forward exchange rate is $F_{to.}$ At time t_0, nothing changes hands.

Obviously, the forward exchange rate F_{to} should be chosen so that the two parties are satisfied with the future settlement, and thus do not ask for any immediate compensating payment This means that the time-t_0 value of a forward contract *concluded* at time to is *zero.* It may, however become positive or negative as time passes and markets move.

In this section, we discuss the structure of this instrument. How do we create a synthetic for an instrument such as this one? How do we decompose a forward contract? Once this is understood, we consider applications of our methodology to hedging, pricing, and risk management.

A general method of engineering a (currency) forward-or, for that matter, any linear instrument - is as follows:

1. Begin with the cash flow diagram in Figure 3-l0 a.
2. Detach and carry the (two) rectangles representing the cash flows into Figure's 3-10b and 3-l0c.
3. Then, add and subtract new cash flows at carefully chosen dates so as to convert the detached cash flows into meaningful financial contracts that players will be willing to buy and sell.
4. As you do this, make sure that when the diagrams are added vertically, the newly added cash flows cancel out and the original cash flows are recovered.

This procedure will become clearer as it is applied to progressively more complicated instru-ments. The first example follows.

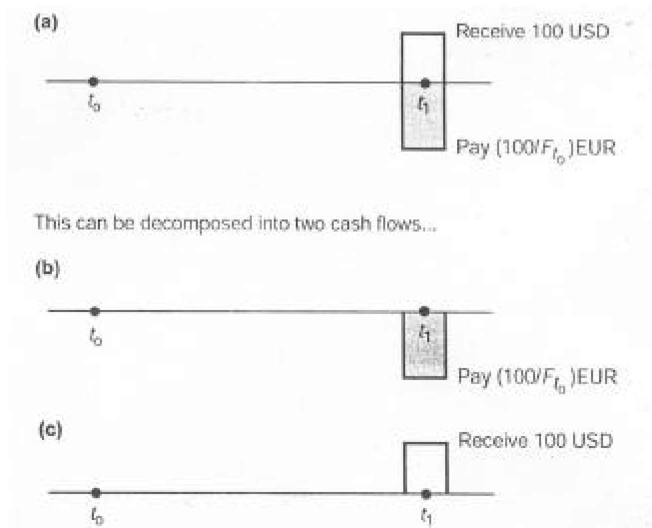


Figure 10 abc

Engineering the Currency Forward

We apply this methodology to engineering a currency forward. The steps are discussed in detail. Our objective is to obtain a contractual equation at the end and, in this way, express the original contract as a *sum* of two or more elementary contracts.

Begin with cash flows in Figure l0a. If we detach the two cash flows, we get Figures -10b and l0c. At this point, nobody would like to buy cash flows in Figure 10b, whereas nobody would sell the cash flows in Figure 10c. Indeed, why pay something without receiving anything in return? So at this point, Figures 10b and 10c cannot represent tradable financial instruments.

However, we can *convert* them into tradable contracts by inserting new cash flows, instep 3 of the methodology. In Figure 10b, we *add* a corresponding cash inflow. In Figure 10c we add a cash *outflow.* By adjusting the size and the timing of these new cash flows, we can turn the transactions in Figures10b and 10c into meaningful financial contracts.

We keep this as simple as possible. For Figure l0b, add a positive cash now, preferably at time t_0.[3] This is shown in Figure l0d. Note that we denote the size of the newly added cash flow by C_{to}^{eur}.

In Figure l0c, add a negative cash flow at time t_0 to obtain Figure 10e. Let this cash now be denoted by C_{to}^{usd} The size of C_{to}^{usd} is not known at this point. except that it has to be in USD.

The *vertical addition* of Figures l0d and l0c should replicate what we started with in Figure 10a. At this point, this *will* not be the case, since C_{to}^{usd} and C_{to}^{eur} do not cancel out at time t_0 as they are denominated in different currencies. But, there is an easy solution to this. The "extra" time *to* cash flows can be eliminated by considering a third component for the synthetic. Consider Figure l0 f where one exchanges C_{to}^{usd} against C_{to}^{eur} at time t_0. After the addition

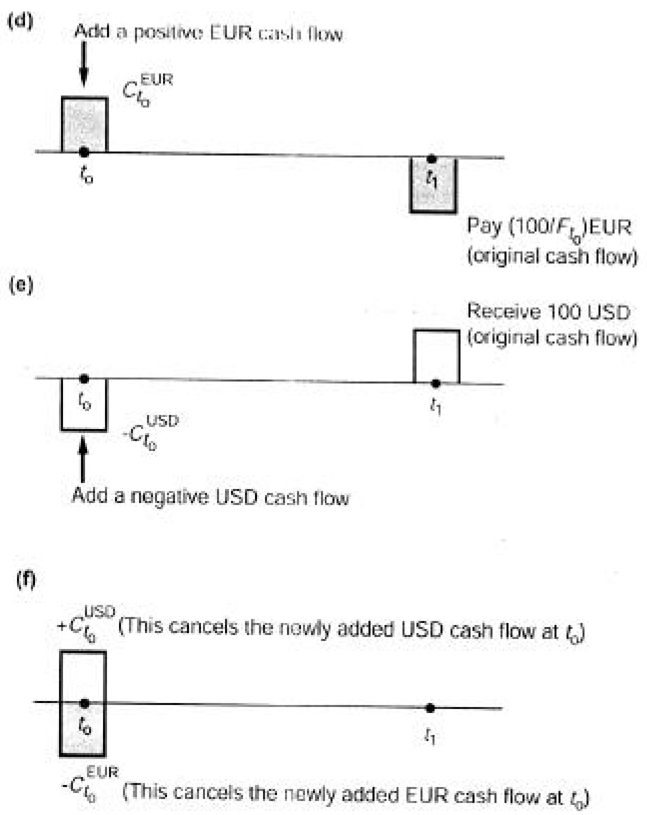


Figure 10def

of this component, a vertical sum *of* the cash flows in Figures 10d, 10e, *and* l0f gives a cash flow pattern identical to the ones in Figure 10a. If the credit risks are the same, we have succeeded in replicating the forward contract with a synthetic.

Which Synthetic?

Yet, it is still not clear what the synthetic in Figures l0d, l0e, and l0f consists of True, by adding the cash flows in these figures we recover the original instrument in Figure l0a, but what kind of contracts do these figures represent? The answer depends on how the synthetic instruments shown in Figures 10d, 10e, and 10f are interpreted.

In fact, these cash flows can be interpreted practically in many different ways. We consider two major cases. The first is a deposit-loan interpretation. The second involves Treasury bills.

A Money Market Synthetic

The first synthetic is obtained using money market instruments. To do this we need a brief review *of* money market instruments. The following lists some important money market instruments, along with the corresponding quote, registration, settlement, and other conventions. The list-is not comprehensive.

Example

Deposits/loans. These mature in less than 1 year. They are denominated in domestic and Eurocurrency units. Settlement is on the same day for domestic deposits and in 2 business' days for Eurocurrency deposits. There is no registration process involved and they are not negotiable.

Certificates of deposit (CD): Generally these mature in up to 1 year. They pay a coupon and are sometimes sold in discount form. They are quoted on a yield basis, and exist both in domestic and Eurocurrency forms; Settlement is on the same day for domestic deposits and in 2 working days for Eurocurrency deposits. They are usually bearer securities and are negotiable.

Treasury bills: These are issued at 13-26-and 52-week maturities. In France, they can also mature in 4 to 7 weeks; in the UK, also in 13 weeks. They are sold on a discount basis (U.S., UK). In other countries, they are quoted on a yield basis. Issued in domestic currency, they are bearer securities and are negotiable.

Commercial paper (CP): Their maturities are 1 to 270 days. They are very short-term securities, issued on a discount basis. The settlement is on the same day, They are bearer securities, are negotiable.

Euro-CP: The maturities range from 2 to 365 days, but most have 30- or 180-day matu-rities. Issued on a discount basis, they are quoted on a yield basis. They can be issued in any Eurocurrency, but in general they are in Eurodollars. Settlement is in 2 business days, they are negotiable.

How can we use these money market instruments to interpret the synthetic for the FX-forward shown in Figure 10?

One money market interpretation is as follows. The cash flow in Figure l0e involves making a payment C_{t0}^{usd} at time t_0 receive USDl00 at a later date. t_1 Clearly, an interbank deposit will generate exactly this cash flow pattern. Then, the C_{t0}^{usd} will he the present value of USD100, where the discount factor can be obtained through the relevant Euro deposit rate.

$$C_{t0}^{usd} = \frac{100}{1 + L_{t0}^{usd}\ (t1 - t2/\ 360)}$$

Note that we are using an *ACT* /360-day basis for the deposit rate L_{t0}^{usd}, since the cash flow is in eurodollars. Also, we are using money market conventions for the interest rate. Given the observed value of L_{to}^{usd}, we can numerically determine the C_{t0}^{usd} by using this equation.

How about the cash flows in Figure l0d? Clearly, this is a *loan* obtained in interbank markets. One receives C_{to}^{eur} at time to, and makes a Euro-denominated payment of $100/\ F_{t0}$ at the later date t_1. The value of this cash flow will be given by

$$C_{t0}^{eur} = \frac{100/F_{t0}}{1 + L_{to}^{eur}(t1 - t2/360)}$$

where the L_{to}^{eur} is the relevant interest rate in Euros.

Finally, we need to interpret the last diagram in 3-10f. These cash flows represent an exchange of C_{t0}^{usd} against C_{to}^{eur} at time t_0. Thus, what we have here is a *spot purchase* of dollars at the rate e_{to}

The synthetic is now fully described:

- Take an interbank loan in Euros (Figure 10d).
- Using these Euro funds, buy spot dollars (Figure 10f).
- Deposit these dollars in the interbank market (Figure l0e).

This portfolio 'would exactly replicate the currency forward, since by adding the cash flows in, Figures l0d, l0e, and 10f, we recover exactly the cash flows generated by a currency forward shown in Figure l0a.

A Synthetic with T-Bills

We can also create a synthetic currency forward using Treasury-bill markets. In fact, let $B(t_0, t_1)usd$ be the time-to price of a default-free discount bond that pays USD100 at time t_1. Similarly, let $B(t_0,t_1)^{eur}$ be the time-t_0 price of a default-free discount bond that pays EUR100 at time t_1. Then the cash flows in Figures l0d, 10e, and l0f can be reinterpreted so as to represent the following transactions:

- Figure 10 is a *short* position in $B(t_o, t_l)^{eur}$ where $1/F_{to}$ units of this security is borrowed and sold at the going market price to generate $B(t_0,t_1)^{eur} / F_{to}$ Euros.
- In Figure l0f, these Euros are exchanged into dollars at the going exchange rate.
- In Figure l0e, the dollars are used to buy one dollar-denominated bond $B(t_0, t_l)^{usd}$.

A time t_1 these operations would amount to exchanging $100/F_{t0}$ EUR against 100USD, given that the corresponding bonds mature at par.

[4.] We remind the reader that if this was a domestic or eurosterling deposit, for example, the day basis would be 365. This is another warning that in financial engineering, conventions matter.

[5.] Disregard for the time being, whether such liquid discount bonds exist in the desired maturities.

Hence, the portfolio

$$\{ Short\ 1/F_{t0}\ units\ of\ B(t_o, t_1)^{eur},\ Long\ B(t_0, t_l)^{usd} \} \quad (3)$$

and the related spot purchase of dollars is another synthetic for the forward currency contract.

Which Synthetic Should One Use?

If synthetics for an instrument can be created in many ways, which one should a financial engineer use in hedging, risk management, and pricing? We briefly comment on this important question.

As a rule, a market practitioner would select the synthetic instrument that is most desirable according to the following attributes: (1) The one that *costs* the least. (2) The one that is most *liquid*, which, *ceteris paribus*, will, in general, be the one that costs the least. (3) The one that is most convenient for *regulatory* purposes. (4) The one that is most appropriate given *balance sheet* considerations. Of course, the final decision will have to be a compromise and will depend on the particular needs of the market practitioner.

Synthetics and Pricing

A major use of synthetic assets is in pricing. Everything else being the same, a replicating portfolio must have the same price as the original instrument. Thus, adding up the value of the constituent assets we can get the *cost* of forming a replicating portfolio. This will give the price of the original instrument once the market practitioner adds a proper margin.

In the present context, *pricing* means obtaining the unknowns in the currency forward, which is the forward exchange rate F_{to} introduced earlier. We would like to determine a set of pricing equations which result in *closed-form* pricing formulas. Let us see how this can be done.

Begin with Figure 10f. This figure implies that the time-to market values of C_{to}^{usd} and' C_{to}^{eur} should be the same. Otherwise, one party will not be willing to go through with the deal. This implies,

$$C_{to}^{usd} = C_{to}^{eur} e_{to} \quad (4)$$

where e_{t0} is the *spot* EUR/USD exchange rate. Replacing from equations (1) and (2):

$$F_{t0} \left(\frac{100}{1 + L_{to}^{usd}(t1,t0/360)} \right) = \left(\frac{100}{1 + L_{to}^{eur}(t1,t0/360)} \right) e_{t0}$$

Solving for the forward exchange rate *Fto*

$$F_{t0} = e_{t0} \left(\frac{1 + L_{to}^{usd}(t1,t0/360)}{1 + L_{to}^{eur}(t1,t0/360)} \right)$$

This is the well-known *covered interest parity* equation. Note that it expresses the "unknown" F_{t0} as a function of variables that can be observed at time *to*. Hence, using the market quotes F_{t0} can be numerically calculated at time t_0 and does not require *any* forecasting effort. [6]

[6] In fact, bringing in a forecasting model to determine the F_{t0} will lead to the wrong market price and may create arbitrage opportunities.

The second synthetic using T-bills give an alternative pricing equation. Since the values evaluated at the current exchange rate, *et*, of the two bond positions needs to be the same, we have

$$F_{t0} B(t_0, t_1)^{usd} = e_{t0} B(t_0, t_1)^{eur} \quad (7)$$

Hence, the F_{t0} priced *off the T-bill markets* will be given by

$$F_{t0} = e_{t0} \frac{B(t_0, t_1)^{eur}}{B(t_0, t_1)^{usd}}$$

If the bond markets in the two currencies are as liquid as the corresponding deposits and loans, the F_{t0} obtained from this synthetic will be very close to the F_{t0} obtained from deposits?

A Contractual Equation

Once an instrument is replicated with a portfolio of other (liquid) assets, we can write a *con-tractual equation* and create a whole sequence of new synthetics. In this section, we will obtain the contractual equation. In the next section, we will show several applications. This section provides a basic approach to constructing static replicating portfolios and hence is central to what will follow.

We have just created a synthetic for currency forwards. The basic idea was that a portfolio consisting of the following instru-ments:

{Loan in EUR, Deposit of USD, spot purchase of USD against EUR}

would generate the same cash flows, at the same *time* periods, with the same *credit risk* as the currency forward. This means that under the (unrealistic) assumptions of

1. No transaction costs
2. No bid-ask spreads
3. No credit risk

we can write the equivalence between the related synthetic and the original instrument as a contractual equation that can conveniently be exploited in practice. In fact, the synthetic using the money market involved three contractual deals that can be summarized by the following "equation":

FX forward Buy USD against EUR	=	Loan Borrow EUR at t_0 for maturity t_1	+	Spot Operation Buy USD against EUR	+	Deposit Deposit USD at t_0 for maturity t_1

This operation can be applied to any two currencies to yield the corresponding FX forward.

The expression shown in Formula (9) is a *contractual equation*. The left-hand side contract leads to the same cash flows generated *jointly* by the contracts on the right-hand side. This does not necessarily mean that the monetary *value* of the two sides is always the same. In fact, one or more of the contracts shown on the right-hand side may not even *exist* in that particular economy and the markets may not even have the opportunity to put a price on them.

Essentially the equation says that the risk-related and cash flow attributes of the two sides are the same. If there is no credit risk, no transaction costs, and if the markets in *all* the involved, instruments are liquid, we expect that *arbitrage* will make the values of the two sides of the contractual equation equal.

Applications

The contractual equation derived earlier and the manipulation of cash flows that led to it may initially be thought of as theoretical constructs with limited practical application. This could not be further from the truth. We now discuss four examples that illustrate how the equation can be skillfully exploited to find solutions to practical, common problems faced by market participants.

Application 1 : A Withholding Tax Problem

We begin with a practical problem of withholding taxes on interest income. Our purpose is not to comment on the taxation aspects but to use this example to motivate the important concept of a synthetic instrument.

The basic idea here is easy to state. If a government imposes withholding taxes on gains from a particular instrument, say a *bond*, and if it is possible to synthetically replicate this instru-ment, then the synthetic may not be subject to withholding taxes. If one learns how to do this, then the net returns offered to clients will be significantly higher- with, essentially, the same risk.

Example

Suppose an economy has imposed a withholding tax on interest income from government bonds. Let this withholding tax rate be 20%. The bonds under question have zero default probability and make no coupon payments. They mature at time- T and their time-t price, is denoted by B(t, T). This means that if

$$B(t, T) = 92 \tag{10}$$

one pays 92 dollars at time t to receive a bond with face value 100. The bond matures at time T, with

$$B(T, T) = 100 \tag{11}$$

Clearly, the interest the bondholder has earned will be given by

$$100 - B(t,T) = 8 \tag{12}$$

But because of the withholding tax, the interest received will only be 6.4 :

$$\text{Interest received} = 8 - .2(8) = 6.4 \tag{13}$$

Thus the bondholder receives significantly less than what he or she earns, especially if we are dealing with a high net worth individual investor. If USD50 million are invested, the bondholder will pay USDO.87 million in withholding taxes. The question is whether a financial engineer can help.

If market professionals can construct a *synthetic* bond that has exactly the same cash flow (and credit risk) characteristics as the original bond except for withholding tax requirements then the problem will be resolved. The synthetic can be constructed so that it is not subject to withholding taxes.[8]

We can immediately use the ideas put forward to form a synthetic for any discount bond using the contractual equation in Formula (9). We discuss this case using two arbitrary currencies called Z and X. Suppose T-bills in both currencies trade actively in their respective markets. The contractual equation written in terms of T-bills gives

FX forward Sell Z against X	=	Short Z – denominate d T-bill	+	Spot Operation Buy currency X with Z	+	Buy X – denominate d T-bill

Manipulating this as an algebraic equation, we can transfer the Z-denominated T-bill to the left-hand side and group another instruments on the right-hand side. After properly changing sides, we obtain

Short Z – denominate d T-bill	= -	FX forward Sell Z against X	+	Spot Operation Buy currency X with Z	+	Buy X – denominate d T-bill

Now, we change the negative signs to positive, which reverses the financial transaction, and obtain a synthetic Z-denominated T-Bill:

Long Z – denominate d T-bill	=	FX forward Buy Z against X	+	Spot Operation Buy currency X with Z	+	Buy X – denominate d T-bill

Thus, in order to construct a synthetic for Z -denominated discount bonds, we first need to use money or T-bill markets of another economy where there is no withholding tax. Let the currency of this country be denoted by the symbol X. According to equation (16) we exchange. Z's into currency X with a spot operation at an exchange rate e_{t0}. Using the X's obtained this, way we buy the relevant X -denominated T-bill. At the same time we forward purchase Z's for time t_1. The geometry of these operations is shown in Figure 3-11. We see that by adding the cash flows generated by the right-hand side operations, we can get exactly the cash flows of a T-bill in Z.

There is a simple logic behind these operations. Investors are taxed on *Z* -denominated bonds. So they use *another* country's markets where there is no withholding tax. They do this in a way that ensures the recovery of the needed *Z's* at time tl by buying *Z* forward. In a nutshell, this is a strategy of carrying funds over time using another currency as a *vehicle* while making sure that the *entry* and *exits* of the position are pinned down at time t_0.

7.2. Application 2: Creating Synthetic Loans

The second application of the contractual equation has already been briefly discussed. Consider the following market event from the year 1997.

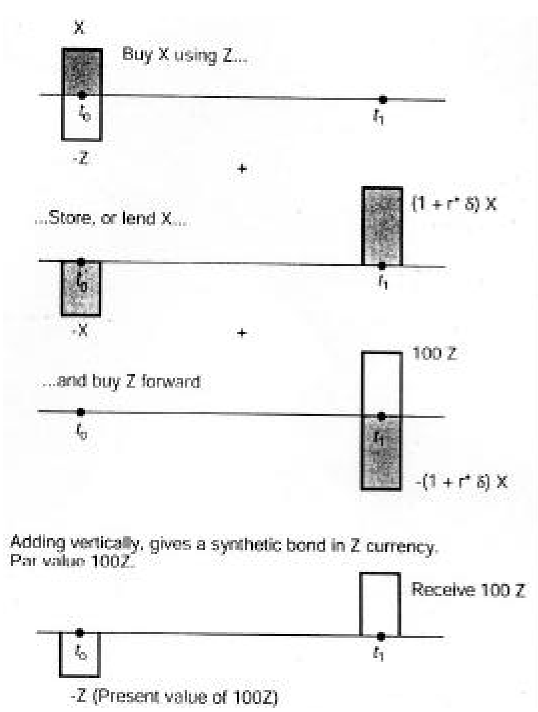


Figure 11

Example

Following the collapse of Hokkaido Takushoku Bank, the" Japanese premium," the extra cost to Japanese banks of raising money in the Eurodollar market increased last week in dramatic style. Japanese banks in the dollar deposit market were said to be pay-ing around 40bp over their comparable U.S. credits, against less than 30bp only a week ago.

Faced with higher dollar funding costs, Japanese banks looked for an alternative source of dollar finance. Borrowing in yen and selling yen against the dollar in the spot market, they bought yen against dollars in the forward market, which in turn caused the U.S. dollar yen forward rate to richen dramatically. (lFR, November 22, 1997)

Readers with no market experience may consider this episode difficult to understand.[9]Yet, the contractual equation in Formula (9) can be used skillfully, to explain the strategy of Japanese banks mentioned in the example. In fact, what Japanese banks were trying to do, was to cre-ate synthetic USD loans. The USD loans were either too expensive or altogether unavailable due to lack of *credit lines.* As such, the excerpt provides an excellent example of a use for synthetics.

We now consider this case in more detail. We begin with the contractual equation in For-mula (9) again, but this time writes it for the USD/JPY exchange rate:

FX forward sell USD against JPY for time t_1	=	Loan Borrow USD with maturity t_1	+	Spot operation Buy JPY pay USD at t_0	+	Deposit Deposit JPY for maturity t_1

Again, we manipulate this like an algebraic equation. Note that, on the right-hand side, there is a loan contract. This is a genuine USD loan, and it can be isolated on the left-hand side by rearranging the right-hand side contracts. The loan would then be expressed in terms of a replicating portfolio.

Loan Borrow USD with maturity t_1		FX forward sell USD against JPY for time t_1		Spot operation Buy JPY pay USD at t_0	-	Deposit Deposit JPY for maturity t_1

Note that because we moved the deposit and the spot opera-tion to the other side of the equality, signs changed. In this context, a deposit with a minus sign would mean reversing the cash flow diagrams and hence it becomes a loan. A spot operation with a minus sign would simply switch the currencies exchanged. Hence, the contractual equation can finally be written as

USD Loan	=	FX Forward Sell USD against JPY for time t_1	+	Spot Operation Buy USD against JPY at t_0	+	A Loan Borrow JPY for maturity t_1

This contractual equation can he used to understand the previous excerpt. According to the quote, Japanese banks that were hindered in their effort to borrow Eurodollars in the interbank (Euro) market instead borrowed Japanese yen in the domestic market, which they used to buy (cash) dollars. But, at the same time, they sold dollars forward against yen in order to hedge their future currency exposure. Briefly, they created exactly the synthetic that the contractual equation implies on the right-hand side. The geometry of these operations is shown in Figure 3-12. Here again we see that the need to obtain funds is carried through by using another currency as an intermediate tool.

Application 3: Capital Controls

Several countries have, at different times, imposed restriction on capital movements. These are known as *capital controls.* Suppose we assume that a spot purchase of USD against the local currency X is *prohibited* in some country.

A financial engineer can construct a *synthetic spot operation* using the contractual relation-ship, since such spot operations were one of the constituents of the contractual equation shown in Formula (9). Rearranging Formula (9), we can write

Spot purchase of USD against X	=	FX forward Sell X against USD for time t_1	+	Loans in USD Borrow USD at t_0	+	Deposit X at t_0 for maturity t_1

The right-hand side will be equivalent to a spot purchase of USD even when there are capital controls. Precursors of such operations were called *parallel loans* and were extensively used by businesses, especially in Brazil and some other emerging markets.[10] The geometry of this situation is shown in Figure 3-13.

Application 4: "Cross" Currencies

Our final example does not make use of the contractual equation in Formula (9) directly. How-ever, it is an interesting application of the notion of contractual equations, and it is appropriate to consider it at this point.

One of the simplest synthetics is the "cross rates" traded in FX markets. A cross currency exchange rate is a price that does not involve USD. The major "crosses" are EUR/JPY, EUR/CHF, GBP/EUR. Other "crosses" are relatively minor. In fact, if a trader wants to purchase Swiss francs in, for example, Taiwan, the trader would carry out two transactions instead of a single spot transaction. He or she would buy U.S. dollars with Taiwan dollars, and then sell the U.S. dollars against the Swiss franc. At the end, Swiss francs are paid by Taiwan dollars. Why would one go through two transactions instead of a direct purchase of Swiss francs in Taiwan? Because it is cheaper to do so, due to lower transaction costs and higher liquidity of the USD/CHF and USD/TWD exchange rates.

We can formulate this as a contractual equation:

Spot purchase of CHF using Taiwan dollars	=	Buy USD against Taiwan dollars	+	Sell USD against Swiss francs

It is easy to see why this contractual equation holds. Consider Figure 3-14. The addition of the cash flows in the top two graphs results in the elimination of the USD element, and one creates a synthetic "contract" of spot purchase of CHF against Taiwan dollars.

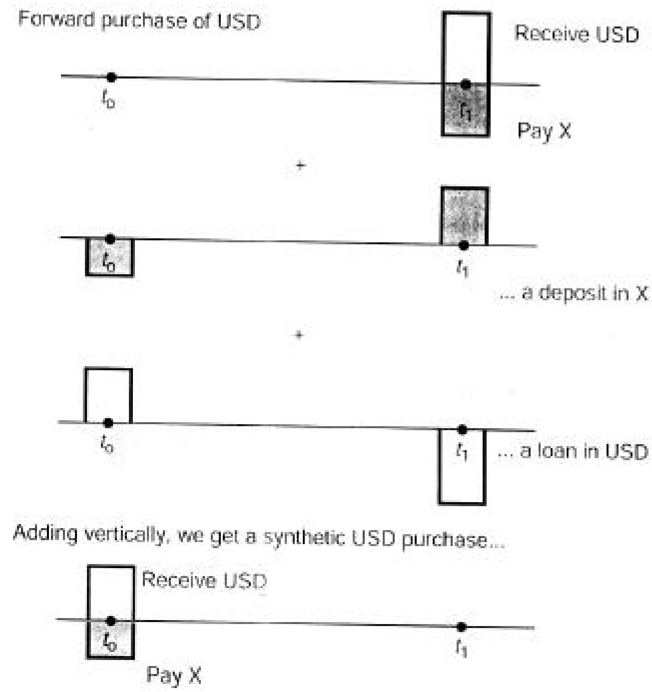


Figure 13

This is an interesting example because it shows that the price differences between the synthetic and the actual contract cannot always be exploited due to transaction costs, liquidity, and other rigidities such as the legal and organizational framework. It is also interesting in this particular case, that it is the synthetic instrument which turns out to be *cheaper.* Thus, before buying and selling an instrument, a trader should always try to see if there is a cheaper synthetic that can do the same job.

A "Better" Synthetic

In the previous sections we created two synthetics for forward FX -contracts. We can now ask the next question: Is there an optimal way of creating a synthetic? Or, more practically, can a trader buy a synthetic cheaply, and sell it to clients after adding a margin, and still post the smallest bid-ask spreads?

FX – Swaps

We can use the so-called *FX-swaps* and pay a single bid-ask spread instead of going through two separate bid-ask spreads as is done in contractual equation (9). The construction of an FX-swap is shown in Figure 15.

According to this figure there are at least two ways of looking at a FX-swap. The FX-swap is made of a money market deposit and a money market loan in different currencies written on the same "ticket". The second interpretation is that we can look at a FX-swap as if the two counterparties *spot* purchase *and forward* sell two currencies against each other, again on the same deal ticket.

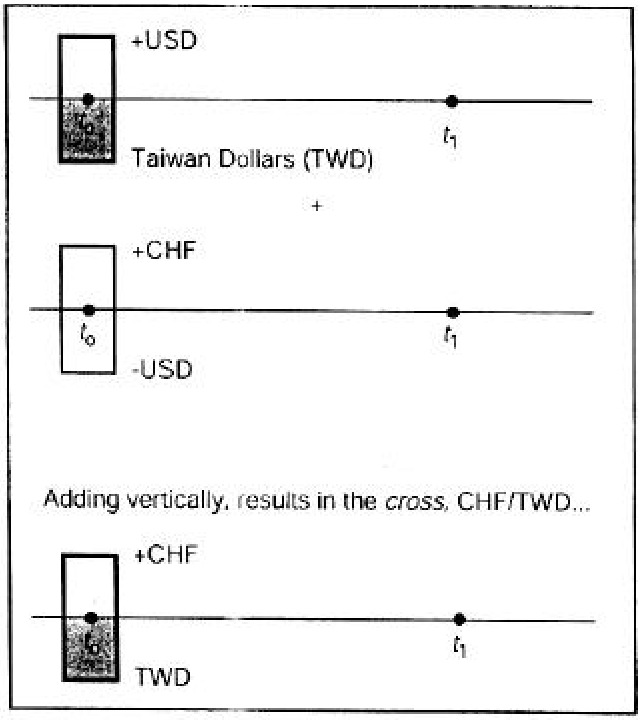


FIGURE 3-14.

as a proportionately higher payment at time t_1' The parties must exchange *different* amounts at time t_1 as compared to the original exchange at t_0.

Advantages

Why would a bank prefer to deal in FX swaps instead of outright forwards? This is an important question from the point of view of financial engineering. It illustrates the advantages of spread products.

FX-swaps have several advantages over the synthetic seen earlier. First of all, FX-swaps are interbank instruments and, normally, are not available to clients. Banks deal with each other every day, and thus will be relatively little counterparty risk in writing such contracts. In liquid markets, the implied bid-ask spread for synthetics constructed using FX-swaps will be smaller than the synthetic constructed from deposits and loans, or T-Bills for that matter.

The second issue is liquidity. How, can a market participant borrow and lend in both curren-cies without moving prices? A FX-swap is again a preferable way of doing this. With a FX-swap, traders are not buying or selling deposits, rather they are exchanging them.

The final advantage of FX-swaps reside in their balance sheet effects, or the lack thereof. The synthetic developed in Figure 3-10 leads to increased assets and liabilities. One borrows new

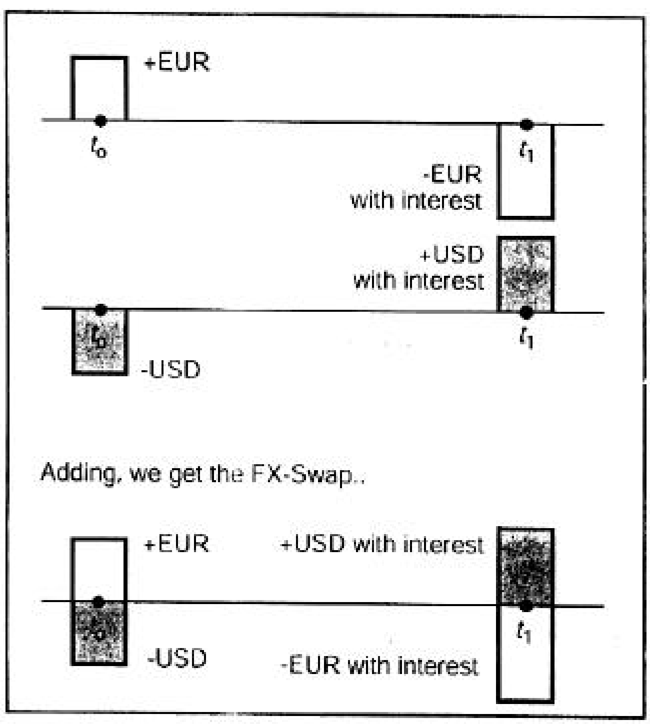


Figure 14 and 15

When combined with a spot operation, FX-swaps duplicate forward currency contracts eas-ily, as seen in Figure 3-16. Because they are swaps of a deposit against a loan, interest rate differentials will play an important role in FX-swaps. After all, one of the parties will be giving away a currency that can earn a higher rate of interest and, as a result, will demand compensation for this “loss.” This compensation will be returned to him or her

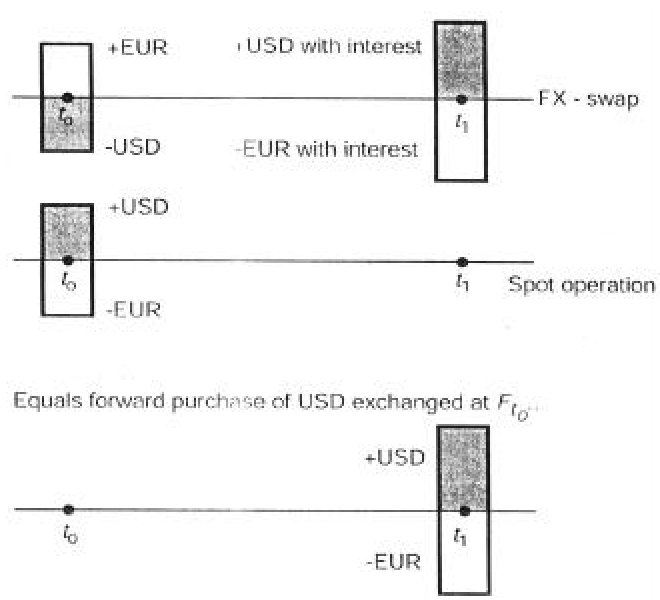


Figure 16

Funds and lends them. Such transactions may lead to new credit risks, new capital re4uirements. FX-swaps are off-balance sheet items, and the synthetic in Figure 16 will have minor balance sheet effects.

Quotation Conventions

Banks prefer to quote *swap* or *forward points* instead of quoting the *outright* forward exchange rate. The relate terminology and conventions are illustrated in the following example:

Example

Suppose outright forward EUR / USD quotes are given by

Bid	Ask
1.0210	1.0220

and that the pot exchange rate quotes are as

Bid	Ask
1.0202	1.0205

Then, instead of the outright forward quotes, traders prefer to quote the forward points obtained by subtracting the corresponding spot rate from the outright forward

Bid	Ask
8	15

In reality forward points are determined directly from equation (6) or (8).

Market conventions sometimes yield interesting information concerning trading activity and the forward FX quotes is a case in point. In fact, there is an important advantage to quoting swap points over the outright forward quotes; this indicates a subtle aspect of market activity. A quote in terms of forward points will essentially be independent of spot exchange rate movements and will depend only on interest rate differentials. An outright forward quote, on the other hand, will depend on the spot exchange rate movements as well. Thus, by quoting forward points, market professionals are essentially *separating* the risks associated with interest rate differentials and spot exchange rate movements respectively. The exchange rate risk will be left to the spot trader. The forward-FX trader will be trading the risk associated with interest rate differentials only.

To see this better, we now look at the details of the argument. Let Ft_0 and et_0 be time-t_1 for-ward and time-to spot exchange rates respectively as given by equation (6). Using the expression in equation (6) and ignoring the bid-ask spreads, we can write approximately,

$$Ft_0 - et_0 \bullet (r^d{}_{t0} - r^f{}_{t0})\left(\frac{t1-t0}{360}\right)e_{t0}$$

where the r^d_{t0}, r^f_{t0} are the relevant interest rates in domestic and foreign currencies, respectively.[11] Taking partial derivatives this equal ion implies that:

$$\partial(Ft_0 - ct_0) \bullet (r^d{}_{t0} - r^f{}_{t0})\left(\frac{t1-t0}{360}\right)$$

If the *daily* movement *of* the spot rate e_{t0} is small, the right hand side will be negligible. In other words, the forward swap quotes would not change *for* normal daily exchange rate movements, if interest rates remain the same and *as long as* exchanges rates are quoted to *four* decimal places. The following example illustrates what this means.

Example

Suppose the relevant interest rates are given by

$$r^d_{t0}, = .03440 \tag{25}$$

$$r^f_{t0} = .02110 \tag{26}$$

where the domestic currency is Euro and the foreign currency is USD. If the EUR / USD exchange rate has a daily volatility of, say, .01 % a day, which is a rather significant move. then, for FX – swaps with 3 months maturity we have the following change in forward points.'

$$\partial(Ft_0 - ct_0) = .01330\left(\frac{90}{360}\right)0.0100$$

which in a market that quotes only four decimal points, will be negligible.

Hence, forward points depend essentially *on* the interest rate differentials, This "separates" exchange rate and interest rate risk and simplifies the work *of* the trader. It also shows that forward FX contracts can be interpreted as if they are "hidden" interest rate contracts.

Futures

Up to this point we considered forward contracts written on currencies only. These are OTC con-tracts, designed according to the needs of the clients and negotiated between two counterparties. They are easy to price and almost costless to purchase.

Futures are different from forward contracts in this respect. Some of the differences are minor; others are more important, leading potentially to significantly different forward and futures prices on the futures prices on the same underlying asset with identical characteristics. Most of these differences come from the design of futures contracts. Futures contracts need to be homogeneous to increase liquidity. The way they expire and the way deliveries are made will be clearly specified, but will still leave some options to the players. Forward contracts are initiated between two specific parties. They can state exactly the delivery and expiration conditions. Futures, on the other hand, will leave some room for last-minute adjustments and these "options" may have market value.

In addition, futures contracts are always marked to market, whereas this is a matter of choice for forwards. Marking to market may significantly alter the implied cash flows and result in some moderate convexities.

To broaden the examination of futures and forwards in this section, we concentrate on commodities at are generally traded via futures contracts in organized exchanges. Let S_t denote the spot price on an underlying commodity and f_t be the futures price quoted in the exchange.

Parameters of a Futures Contract

We consider two contracts in order to review the main parameters involved in the design of a futures. The key point is that most aspects of the transaction need to be pinned down to make a homogenous and liquid contract. This is relatively easy and straightforward to accomplish in the case of a relatively standard commodity such as soybeans.

EXAMPLE: CBO T Soybeans Futures

1. **Contract size:** 5000 bushels.
2. **Deliverable grades:** No.2 yellow at par, No.1 yellow at 6 cents per bushel over contract price, and No.3 yellow at 6 cents per bushel under contract price. (Note that in case a

trader accepts the delivery, a special type of soybeans will be delivered to him or her. The trader may, in fact, procure the same quantity under better conditions from someone else. Hence, with a large majority of cases, futures contracts do not end with delivery. Instead, the position is unwound with an opposite transaction sometime before expiration.)

3. Tick size: quarter-cent/bushel ($12.50/contract).
4. Price quote: Cents and quarter-cent/bushel.
5. Contract months: September, November, January, March, May, July, and August. (Clearly, if the purpose behind a futures transaction is delivery, then forward. contracts with more flexible delivery dates will be more convenient.),
6. Last trading day: The business day prior to the 15th calendar day of the contract, month.
7. Last delivery day: Second business day following the last trading day of the delivery month.
8. Trading hours: Open outcry: 9:30 a.m. to 1:15 p.m. Chicago time, Monday through Friday. Electronic, 8:30 p.m. to 6:00 a.m. Chicago time, Sunday through Friday. Trading in experience contracts closes at noon on the last trading day.
9. Daily price limit: 50 cents/bushel ($2500/contract) above or below the previous day's settlement price. No limit in the spot month (Limits are lifted two business days before the spot month begins).

A second example is from financial futures. Interest rate futures constitute some of the most liquid instruments in all markets. They are, again, homogenized contracts and will be discussed in the next chapter.

EXAMPLE: LIFFE 3-Month Euro Libor Interest-Rate Futures

1. Unit of trading: Euro 1,000,000.
2. Delivery months: March, June, September, and December. June 2003 is the last contract month available for trading.
3. Price quotes: 100 minus rate of interest. (Note that prices are quoted to three decimal places. This means that the British Bankers Association (BBA) Libor will be rounded to three decimal places and will be used in settling the final value of the contract.)
4. Minimum price movement: (Tick size and value) 0.005(12.50).
5. Last trading day: Two business days prior to the third Wednesday of the delivery month.
6. Delivery day: First business day after the last trading day.
7. Trading hours: 07:00 to 18:00.

Such Eurocurrency futures contracts will be discussed in the next chapter and will be revisited several times later. In particular, one aspect of the contract that has not been listed among the parameters noted here has interesting financial engineering implications. Eurocurrency futures have a quotation convention that implies a linear relationship between the forward interest rate and the price of the futures contract. This is another example of the fact that conventions are indeed important in finding the right solution to a financial engineering problem.

One final, but important point. The parameters of futures contracts are sometimes revised by Exchanges; hence the reader should consider the information provided here simply as examples and check the actual contract for specifications.

Marking to Market

We consider the cash flows generated by a futures contract and compare them with the cash flows on a forward contract on the same underlying. It turns out that, unlike forwards, the effective maturity of a futures position is, in fact, 1 day. This is due to the existence of marking to market in futures trading. The position will be marked to market in the sense that every night the exchange will, in effect, close the position and then reopen it at the new settlement price. It is best to look at this with a precise example. Suppose a futures contract is written on one unit of a commodity with spot price S_t Suppose t is a Monday and that the expiration of the contract is within 3 trading days:

$$T = t+3 \qquad (27)$$

Suppose further that during these days, the *settlement prices* follow the trajectory

$$f_t > f_t +1 > f_t +2 = f_t + 3 \qquad (28)$$

What cash flows will be generated by a long position in one futures contract if at expiration date *T* the position is closed by taking the offsetting position?[12]

The answer is shown in Figure 3-1 7. Marking to market is equivalent to forcing the long (short) position holder to close his position at that day's settlement price and reopen it again at

[12] Instead of taking the offsetting position and canceling out any obligations with respect to the clearinghouse, the trader could chose to accept delivery.

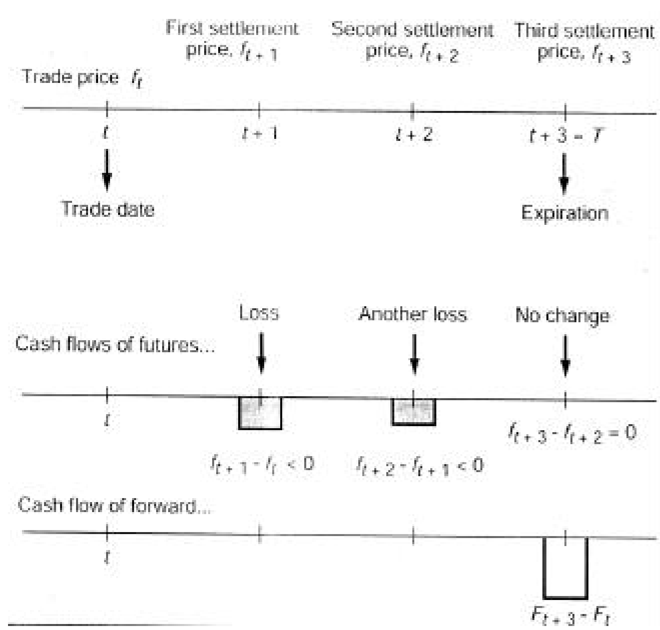


Figure 17

the same price. Thus, at the end of the first trading day after the trade, the futures contract that was "purchased" at f*t* will be "sold" at the f*1*$+_1$ shown in (28) for a loss:

$$f_t +_1 - f_t > 0 \qquad (29)$$

Similarly, at the end of the second trading day, marking to market will lead to another loss:

$$f_{t+2} - f_{t+1} > 0 \quad (30)$$

This is the case since, according to trajectory in (28), prices have declined again. The expiration date will see no further losses, since, by chance, the final settlement price is the same as the previous day's settlement.

In contrast, the last portion of Figure 17 shows the cash flows generated by the forward prices F_t. Since there is no marking to market (in this case), the only capital loss occurs at the expiration of the contract. Clearly, this is a very different cash flow pattern.

Cost of Carry and Synthetic Commodities

What is the carry cost of a position? We will answer this question indirectly. In fact, ignoring the mark to market and other minor complications, we first apply the contractual equation developed earlier to create synthetic commodities.

For example, suppose St represents spot coffee, which is the underlying asset for a futures contract with price f_t and expiration date T, $t_o < T$. How can we create a synthetic for this contract? The answer is quite similar to the case of currencies. Using the same logic, we can , write a contractual equation:

Long coffee futures expiration *T*	=	A Loan Borrow USD at to for maturity *T*	+	Spot operation Buy 1 unit of spot coffee for S_{to}	+	Store the coffee at a cost q_{to} a day until *T*

-We can use this equation to obtain two results. First, by rearranging the contracts, we create a synthetic spot:

Spot operation Buy one unit of spot coffee for S_{to}	= -	A Loan Borrow USD at to for maturity *T*	+	Long coffee futures Expiration *T*	-	Store the coffee at a cost q_{to} a day until *T*

In other words after changing signs, we need to borrow one unit of coffee, make a deposit *of* S_{to} dollars, and go long a coffee futures contract. This will yield a synthetic spot.

Second, the contractual equation can be used in pricing. In fact, the contractual equation gives the carry cost *of* a position. To see this first note that according to equation (31) the value of the synthetic the same as the value of the original contract. Then, we must have

$$f_{to} = (1 + r_{to}\, d\,) S_{to} + q_{to} (T - t_o) \quad 33)$$

where d is the factor of days' adjustment to the interest rate denoted by the symbol $r_{to.}$.

If storage costs are expressed as a percentage *of* the price, at an annual rate, just like the interest rates, this formula becomes

$$f_{to} = (\,1 + r_{to}\, d\, + q_{to}\, d\,)\, S_{to} \quad (34)$$

According to this, the more distant the expiration of the contracts is, the higher its price. This means that futures term structures would normally be upward sloping as shown in Figure 3-18. Such curves are said to be in contango. For some commodities, storage is either not possible (e.g., due to seasons) or prohibitive (e.g., crude oil). The curve may then have a negative slope and is said to be in backwardation. Carry cost is the interest plus storage costs here.

A Final Remark

There are no upfront payments but buying futures or forward contracts is not costless. Ignoring any guarantees or margins that may be required for taking futures positions, taking forward or futures positions does involve a cost. Suppose we consider a storable commodity with spot price

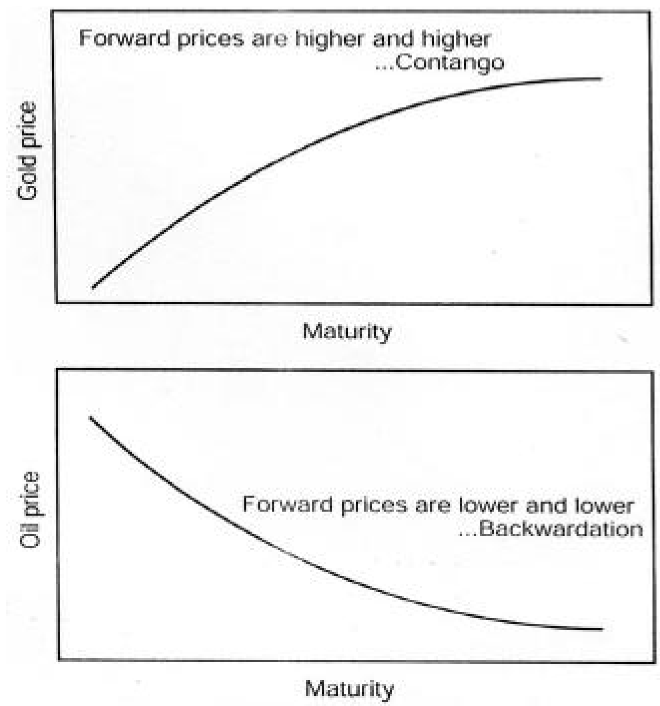


Figure 18

Pto. Let the forward price be denoted by $P^f_{to.}$ Finally, suppose storage costs and all such effects are zero. Then the futures price is given by

where the r_{to} is the appropriate interest rate that applies for the trader, and where ä is the time to expiration as a proportion of a year.

Now, suppose the spot price remains the *same* during the life of the contract. This means that the difference

$$P^f_{to} - P_{to} = r_{to}\, d\, P_{to} \quad (36)$$

is the cost of taking this position. Note that this is as if we had borrowed P_{to} dollars for a "period" d in order to carry a long position. Yet there has been no exchange of principals. In the case of a default, no principal will be lost.

Conventions for Forwards

Forwards in foreign currencies have special quotation conventions. As mentioned earlier, in discussing FX-swaps, markets do not quote outright forward rates, but the so-called forward points. These are the difference between the forward rate found using the pricing equation in Formula (22) and the spot exchange rate:

$F_{to} - e_{to}$ (37)

They are also called "pips" and written as bid/ask. We give an example for the way forward points are quoted and used.

Example

Suppose the spot and forward rate quotes are as follows:

EUR/USD	Bid	Ask
Spot	0.8576	0.8572
1yr	-28.3	-27.3
2yr	44.00	54.00

From this table we can calculate the outright forward exchange rate F_{to}.

For year 1, subtract the negative pips in order to get the outright forward rates:

$$\left(\frac{44}{0.8567+10000}\right)/\left(\frac{54}{0.8572+10000}\right) \quad (39)$$

Forward points give the amount needed to adjust the spot rate in order to obtain the outright, forward exchange rate. Depending on the market, they are either added to or subtracted from the spot exchange rate. We should discuss briefly some related conventions.

There are two cases of interest. First, suppose we are given the following forward point quotes (second row) and spot rate quotes (first row) for EUR/USD:

Bid	Ask
1.0110	1.0120
12	16

Next note that the forward point listed in the "bid" column is lower than the forward point listed in the "ask" column. If forward point quotes are presented this way, then the points will be added to the last two digits of the corresponding spot rate.

Thus, we will obtain

Bid forward outright =1.0110 + .0012 = 1.0122 (40)

Ask forward outright =1.0120 + .0016 = 1.0136 (41)

Note that the bid-ask spread on the forward outright will be greater than the bid-ask spread on the spot.

Second, suppose, we have till' following)'. quotes:

Bid	Ask
1.110	1.0120
23	18

Here the situation is reversed. The forward point listed in the "bid" column is greater than the forward point listed in the "ask" column. Under these conditions, the forward points will be subtracted from the last two digits of the corresponding spot rate. Thus, we will obtain

Bid forward outright =1.0110 - .0023 = 1.0087

Ask forward outright =1.0120 - .0018 = 1.0102

Note that the bid-ask spread on the forward outright will again be greater than the bid-ask spread on the spot. This second case is due to the fact that sometimes minus sign is ignored in quotations of forward points.

Exercises

1. On March 3, 2000, the Financial Accounting Standards Board, a crucial player in financial engineering problems, published a series of important new proposals concerning the accounting of certain derivatives. It is known as Statement 133 and affects the daily lives of risk managers and financial engineers significantly. One of the treasurers who is affected by the new rules had the following comment on these new rules:

   Statement 133 in and of itself will make it a problem from an accounting poInt of view to do swaps. The amendment does not allow for a distinction to be made. between users of aggressive swap hedges and those involved in more typical, swaps. According to IFR this treasurer has used synthetic swaps to get around [the FAS 133].[13].

   a. Ignoring the details of swaps as an instrument, what is the main point in FAS 133 that disturbs this market participant?

   b. How does the treasurer expect to get around this problem by constructing synthetics?

2. In this question we consider a gold miner's hedging activities.

   a. What is the natural position of a gold miner? Describe using payoff diagrams.

   b. How would a gold miner hedge her position if gold prices are expected to dropsteadily over the years? Show using payoff diagrams.

   c. Would this hedge ever lead to losses?

3. Today is March 1,2004. The day-count basis is actual/365. You have the following contracts on your FX-book.

CONTRACT A: On March 15, 2004, you will sell 1,000,000 EUR at a price F^1_t dollars per EUR.

CONTRACT B: On April 30, 2004, you will buy 1,000,000 EUR at a price F^2_t dollars per EUR.

a. Construct one synthetic equivalent of each contract.

b. Suppose the spot EUR/USD is 1.1500/1.1505. The USD interest rates for loans under 1 year equal 2.25/2.27, and the German equivalents equal 2.35/2.36. Calculate the F^i_t numerically.

c. Suppose the forward points for F^1_t that we observe in the markets is equal to 10/20. How can an arbitrage portfolio be formed?

4. Consider the following instruments and the corresponding quotes. Rank these instruments in increasing order of their yields.

Instrument	Quote
30-day U.S.T-bill	5.5
30-day U.K.T-bill	5.4
30-day ECP	5.2
30-day interbank deposit US	5.5
30-day US CP	5.6

a. You purchase a ECP (Euro) with the following characteristics

Value Date	July 29, 2002
Maturity	September 29, 2002
Yield	3.2%
Amount	10,000,000 USD

What payment do you have to make?

ENGINEERING OF INTEREST RATE DERIVATIVES

Objectives

- Completion of this lesson will help the students about the financial engineering application that use forward rate loans and FRAs.

Introduction

Foreign currency and commodity forwards are the simplest types of derivative instruments. The instruments described in this chapter are somewhat more complicated. The chapter discusses financial engineering methods that use *forward loans, Eurocurrency futures,* and *forward rate agreements* (FRAs). The discussion prepares the ground for the next two chapters on swap-based financial engineering. In fact, the FRA contracts considered here can be regarded as precursors of plain vanilla swaps.

Interest rate strategies and risk management present more difficulties than FX, equity, or commodity-related instruments for many reasons. One can mention two. First of all, the payoff of an interest rate derivative depends, by definition, on some interest rate(s). In order to price the instrument, one needs to apply discount factors to the future payoffs and calculate the relevant present values. But the discount factor itself is an interest rate-dependent concept. If interest rates are stochastic, the present value of an interest rate – dependent cash flow will be a nonlinear random variable; the resulting expectations may not be as easy to calculate. There will be two sources of any future fluctuations – those due to future cash flows themselves and those due to changes in the discount factor applied to these cash flows. When dealing with equity or commodity derivatives, such nonlinearities are either not present or have a relatively minor impact on pricing.

Second, every interest rate is associated with a maturity or tenor. This means that, in case of interest rate derivatives we are not dealing with a single random variable, but with vector-valued stochastic processes. The existence of such a vector – valued random variable requires new methods of pricing, risk management, and strategic position taking.

A Convergence Trade

Before we start discussing replication of elementary interest rate derivatives we consider a real life example.

For a number of years before the new European currency was born, there was significant uncertainty as to which countries would be permitted to form the group of Euro users. During this period, market practitioners put in place the so-called convergence plays. The reading that follows is one example.

Example

Last week traders took positions on convergence at the periphery of Europe.Traders sold the spread between the Italian and Spanish curves. JP Morgan urged its customers to buy a 12x24 Spanish forward rate agreement (FRA) and sell a 12x24 Italian FRA. According to the bank, the spread, which traded at 133bp, would move down to below 50bp.

The logic of these trades was that if Spain entered the single currency, then Italy would also do so. Recently, the Spanish curve has traced below the Italian curve. According to this logic, the Italian yield curve would converge on the Spanish yield curve, and traders would gain. (Episode based on IFR issue number 1887).

In this episode, traders buy and sell spreads in order to benefit from a likely occurrence of an event. These spreads are bought and sold using the FRAs, which we discuss in this chapter. If the two currencies converge, the difference between Italian and Spanish interest rates will decline. The FRA positions will benefit. Note that market professionals call this selling the spread. As the spread goes down, they will profit hence, in a sense they are short the spread.

This lesson develops the financial engineering methods that use forward loans, FRAs, and Eurocurrency futures. We first discuss these instruments and obtain contractual equations that can be manipulated usefully to produce other synthetics. The synthetics are used to provide pricing formulas.

Libor and Other Benchmarks

We first need to define the concept of Libor rates. The existence of such reliable benchmarks is essential for engineering interest rate instruments.

Libor is an arithmetic average interest rate that measures the cost of borrowing from the point of view of a panel of preselected banks in London. Libor interest rates are published daily at 11:00 London time for nine currencies.

Euribor is a similar concept determined in Brussels by polling a panel of banks in continental Europe. These two benchmarks will obviously be quite similar. London banks and Frankfut banks face similar risks and similar costs of funding. Hence they will lend euros at approximately the same rate. But Libor and Euribor may have some slight differences due to the composition of the panels used.

Important Libor maturities are overnight, one week, one, two, three, six, nine, and twelve months. A plot of Libor rates against their maturities is called the Libor curve.

Libor is a money market yield and in most curr3encies it is quoted on the ACT/360 basis. Derivatives written on Libor are called Libor instruments. Using these derivatives and the underlying Euromarkets loans, banks create Libor exposure. Tibor (Tokyo) and Hibor (Hong Kong) are examples of other benchmarks that are used for the same purpose.

When we use the term "interest rates" in this chapter, we mean Libor rates. We can now define the major instruments that will be used. The first of these are the forward loans. These are not liquid, but they make a good starting point. We then move to forward rate agreements and to Eurocurrency futures.

Forward Loans

A forward loan is engineered like any forward contract, except that what is being bought or sold is not a currency or commodity, but instead, a loan. At time t_0 we write a contract that will settle at a future date t_1. The trader receives (delivers) a loan that matures at t_2, $t_1<t_2$. The contract will specify the interest rate that will apply to this loan. This interest rate is called the forward rate and will be denoted by F (t_0, t_1, t_2). The forward rate is determined at t_0. The t_1 is the start date of the future loan, and t_2 is the date at which the loan matures.

The situation is depicted in Figure 4-1. We write a contract at t_0 such that at a future date, t_1, USD100 are received; the principal and interest are paid at t_2. The interest is $F_{t0}\, d$, where d is the day-count adjustment:

$$d = \frac{t2 - t1}{360}$$

To simplify the notation, we abbreviate the F (t_0, t_1, t_2) as F_{t0}. Like in Chapter 3, the day – count convention needs to be adjusted if a year is defined as having 365 days.

Forward loans permit a great deal of flexibility in balance sheet, tax, and risk management. The need for forward loans arises under the following conditions,

- A business would like to lock in the "current" low borrowing rates from money markets.
- A bank would like to lock in the "current" high lending rates.
- A business may face a floating – rate liability at time t_1. The business may want to hedge this liability by securing a future loan with a known cost.

It is straightforward to see how forward loans help to accomplish these goals. With the forward loan of figure 4-1, the party has agreed to receive 100 dollars at t_1 and to pay them back at t_2 with interest. The key point is that the

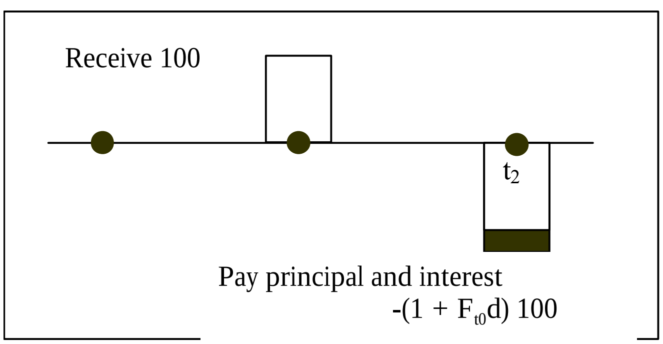


Figure 1.

Interest rate on this forward loan is fixed at time t_0. The forward rate F (t_0, t_1, t_2) "locks in" an unknown future variable at time t_0 and thus eliminates the risk associated with the unknown rate. The L_{t1} is the Libor interest rate for a $(t_{-2}\text{-}t_1)$ period loan and can be observed only at the future date t1. Fixing F (t_0, t_1, t_2) will eliminate the risk associated with L_{t1}.

The lesson discusses several examples involving the use of forward loans and their more recent counterparts, forward rate agreements.

Replication of a Forward Loan

In this section we apply the techniques developed in Chapter 3 to forward loans and thereby obtain synthetics for this instrument. More than the synthetic itself, we are concerned with the methodology used in creating it. Although forward loans are not liquid and rarely traded in the markets, the synthetic will generate a contractual equation that will be useful for developing contractual equations for FRAs, and the latter are liquid instruments.

We first decompose the forward loan cash flows into separate diagrams and then try to convert these into known liquid instruments by adding and subtracting appropriate new cash flows. This is done so that, when added together, the extra cash flows cancel each other out and the original instrument is recovered. Figure 2 displays the following steps:

We begin with cash flow diagram for the forward loan shown in figure 4-2a. We detach the two cash flows into separate diagrams. Note that at this stage, these cash flows cannot form tradeable contracts. Nobody would want to buy 4-2c, and everybody would want to have 2b.

We need to transform these cash flows into tradeable contracts by adding compensating cash flows in each case. In figure 4-2b we add a negative cash flow, preferably at time t_0[3]. This is shown in figure 4-2d. Denote the size of the cash flow by $-C_{t0}$.

In Figure 2c, add a positive cash flow at time t_0, to obtain Figure 2e. The cash flow has size $+C_{t0}$.

Make sure that the vertical addition of Figures 2d and 2e will replicate what we started with in figure 2a. For this to be the case, the two newly added cash flows have to be identical in absolute value but different in sign. A vertical addition of Figures 2d and 2e will cancel any cash exchange at time t_0, and this is exactly what is needed to duplicate figure 2a.

At this point, the cash flows of figure 2d and 2e need to be interpreted as specific financial contracts so that the components of the synthetic can be identified. There are many ways to do this. Depending on the interpretation the synthetic will be constructed using different assets.

Bond Market Replication

A first synthetic can be obtained using bond and T-bill markets. Although this is not the way preferred by practitioners, we will see that the logic is fundamental to financial engineering.

Suppose default-free pure discount bonds of specific maturities denoted by $\{B (t_0, t_i), I = 1, \ldots. n\}$ trade actively. They have par value of 100.

Then, within the context of a pure discount bond market, we can interpret the cash flows in Figure 2d as a long position in the t_1-maturity discount bond. The trader is paying C_{t0} at time t_0 and receiving 100 at t_1. This means that

$$B (t_0, t_1) = C_{t0} \qquad (2)$$

Hence, the value of C_{t0} can be determined if the bond price is known.

The synthetic for the forward loan will be fully described once we put a label on the cash flows in Figure 2e. What do these cash flows represent? These cash flows look like an appropriate short position in a t_2 maturity discount bond.

Does this mean we need to short one unit of the B (t_0, t_2)? The answer is no, since the time t_0 cash flow in Figure 4-2e has to equal C_{t0}[5]. however, we know that

$$B(t_0, t_2) < B(t_0, t_1) = C_{t0} \quad (3)$$

A t_2-maturity bond will necessarily be cheaper than a t_1-maturity discount bond. Thus, shorting one t_2-maturity discount bond will not generate sufficient time – t_0 funding for the position in Figure 4-2d. The problem can easily be resolved, however, by shorting not one but ? bonds such that

$$? \; B(t_0, t_1) = C_{t0} \quad (4)$$

But we already know that B (t_0, t_1) = C_{t0}. So the ? can be determined easily:

$$? = \frac{B(t_0, t_1)}{B(t_0, t_2)} \quad (5)$$

According to (3) ? will be greater than one. This particular short position will generate enough cash for the long position in the t_1 maturity bond. Thus, we finalized the first synthetic for the forward loan:

{Buy one t_1-discount bond, short $\frac{B(t_0, t_1)}{B(t_0, t_2)}$ units of the t_2-discount bond} (6)

To double – check this result, we add Figures 2d and 2e vertically and recover the original cash flow for the forward loan in figure 2a.

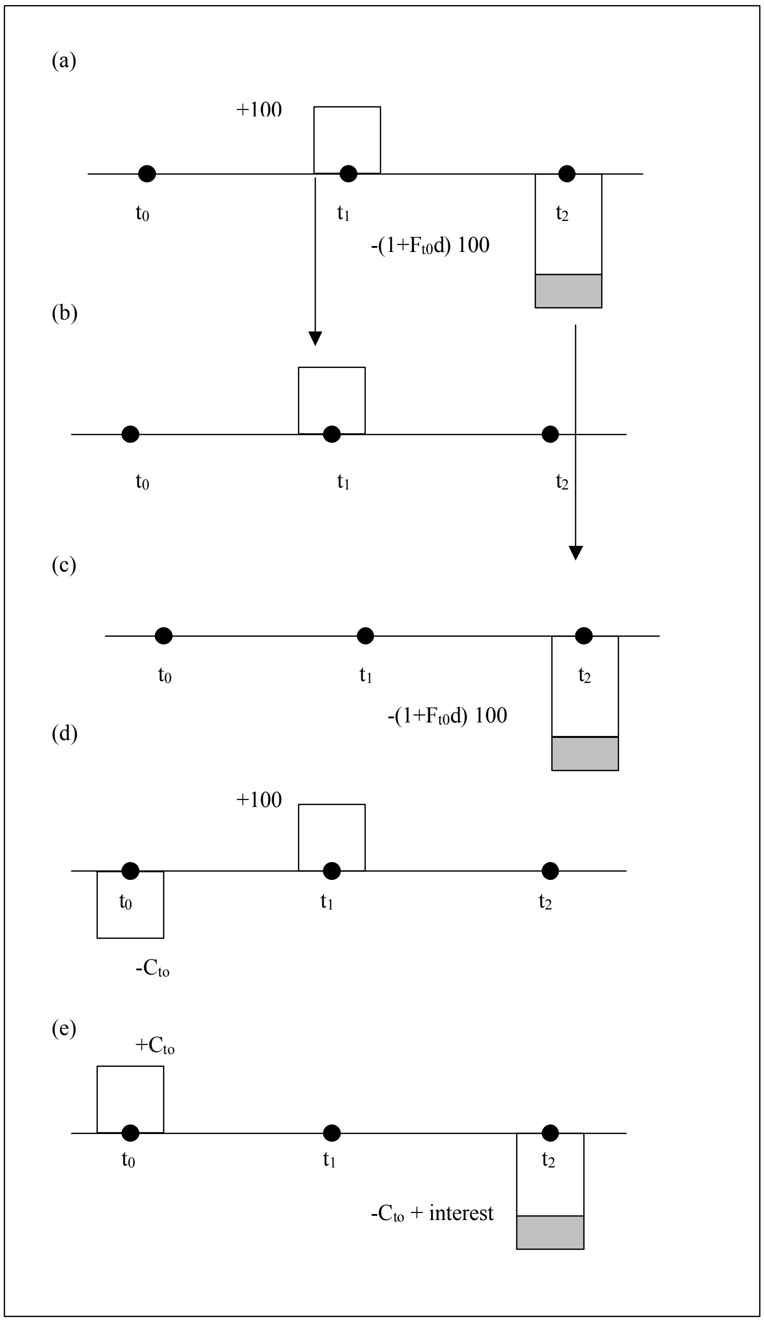


Figure 2

Pricing

If markets are liquid and there are no other transaction costs, arbitrage activity will make sure that the cash flows from the forward loan and from the replicating portfolio (synthetic) are the same. In other words the sizes of the times-t_2 cash flows in figures 4-2a and 4-2e should be equal. This implies that

$$1+F\,(t_0, t_1, t_2)d = \frac{B\,(t_0, t_1)}{B\,(t_0, t_2)}$$

where the d is , as usual, the day-count adjustment:

$$d = \frac{t_2 - t_1}{360} \tag{8}$$

Obviously, day-count parameter needs to be adjusted if the convention is 365 days. This arbitrage relationship is of fundamental importance in financial engineering. Given liquid bond prices $\{B(t_0, t_1), B(t_0, t_2)$, we can price the forward loan the off the bond markets using this equation. More important, equality (7) shows that there is a crucial relationship between forward rates at different maturities and discount bond prices. But discount bond prices are discounts which can be used in obtaining the present values of future cash flows. This means that forward rates are of primary importance in pricing and risk managing financial securities.

Before we consider a second synthetic for the forward loan, we prefer to discuss how all this relate to the notion of arbitrage.

Arbitrage

In fact, suppose Equality (7) does not hold. What happen when the equality in formula (7) breaks down? We analyze two cases assuming that are no bid-ask spreads. First, suppose market quotes at time t_0 are such that

$$(1+F_{t0}d) > \frac{B\,(t_0, t_1)}{B\,(t_0, t_2)} \tag{9}$$

where the forward rate $F(t_0, t_1, t_2)$ is again abbreviated as F_{t0}. Under these conditions, a market participant can secure a synthetic forward loan in bond markets. This will guarantee positive arbitrage gains. This is the case since the "synthetic" *funding cost,* denoted by F^*_{t0},

$$F^*_{t0} = \frac{B\,(t_0, t_1)}{dB(t_0, t_1)} - \frac{1}{d}$$

Will be less than the forward rate, F_{t0}. the position will be risk less if it is held until maturity date t_2.

These arbitrage gains can be secured by (1) shorting $\frac{B\,(t_0, t_1)}{B\,(t_0, t_2)}$ units of the t_2-bond, which generates $B\,(t_0, t_1)$ dollars at time t_0, then(2) using these funds buying one t_1 – maturity bond, and (3) at time t_1 lending, at time t_2, the trader would owe $\frac{B\,(t_0, t_1)}{B\,(t_0, t_2)}$ 100 and would receive $(1+F_{t0}d)100$. The latter amount is greater, given the condition (9).

Now consider the second case. Suppose time – t_0 markets quote:

$$(1+F_{t0}d) < \frac{B\,(t_0, t_1)}{B\,(t_0, t_2)} \tag{11}$$

Then, one can take the reverse position. Buy $\frac{B(t_0,t_1)}{B(t_0,t_2)}$ units of the t_2 – bond $\frac{B(t_0,t_1)}{B(t_0,t_2)}$ at time t_0. To fund this short a $B\,(t_0, t_1)$ bond and borrow 100 forward. When time t_2 arrives, receive the 100 and pay off the forward loan. This strategy can yield arbitrage profits since the funding cost during $[t_1, t_2]$ is lower than the return.

Money Market Replication

Now assume that all maturities of deposits up to 1 year are quoted actively in the interbank money market. Also assume there are no arbitrage opportunities. Figure 3 shows how an alternative synthetic can be created. The cash flows of a forward loan are replicated in Figure 3a.

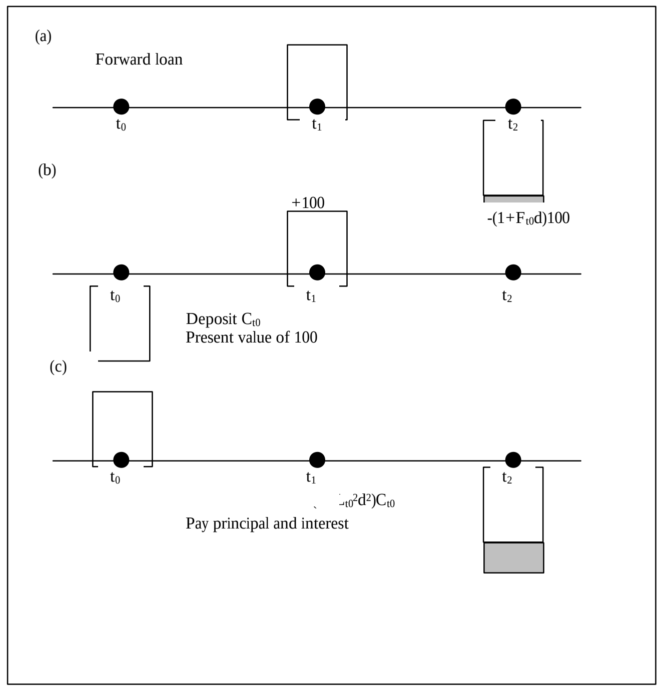


Figure 3

Figure 3c shows a Euromarkets loan. Ct_0 is borrowed at the inter-bank rate L^2_{t0}[6]. The time $-t_2$ cash flow in Figure 3c needs to be discounted using this rate. This gives

$$C_{t0} = \frac{100(1+F_{t0}d)}{(1+L^2_{t0}d^2)} \tag{12}$$

Where $d^2 = (t_2 - t_0)/360$

Then, C_{t0} is immediately re-deposited at the rate L^1_{t0} at the shorter maturity. To obtain

$$C_{t0}\,(1+L^1_{t0}d^1) = 100 \tag{13}$$

With $d^1 = (t_1 - t_0)/360$. This is shown in figure 4-3b

Adding Figures 3b and 4c vertically, we again recover the cash flows of the forward loan.Thus, the two Euro-deposits form a second synthetic for the forward loan.

Pricing

We can obtain another pricing equation using the money market replication. In figure 3, if the credit risks are the same, the cash flows at time t_2 would be equal, as implied by equation (12). This can be written as

$$(1+F_{t0}{}^{d})100 = C_{t0}(1+L^{2}{}_{t0}d^{2}) \qquad (14)$$

where $d = (t_2 - t_1)/360$. We can substitute further from Formula (13) to get the final pricing formula:

$$(1+F_{t0}{}^{d})100 = \frac{100(1+L^{2}{}_{t0}d^{2})}{100(1+L^{1}{}_{t0}d^{1})} \qquad (15)$$

Simplifying,

$$(1+F_{t0}{}^{d}) = \frac{(1+L^{2}{}_{t0}d^{2})}{(1+L^{1}{}_{t0}d^{1})} \qquad (16)$$

This formula prices the forward loan off the money markets. The formula also shows the important role played by *Libor* interest rates in determining the forward rates.

Contractual Equations

We can turn these results into analytical contractual equations. Using the bond market replication. We obtain

Forward loan that begins at t_1 and ends at t_2	Short $B(t_0, t_1)/B(t_0, t_2)$ units of t_2 maturity bond	Long a t_1 – maturity bond.	(17)

If we use the money markets to construct the synthetic, the contractual equation becomes

Forward loan that begins at t_1 and ends at t_2	=	Loan with maturity t_2	+	Deposit with maturity t_1.	(18)

These contractual equations can be exploited can be exploited for finding solutions to some routine problems encountered in financial markets although they do have drawbacks. Ignoring these for the time being we give some examples.

Applications

Once a contractual equation for a forward loan is obtained, it can be algebraically manipulated as in chapter 3, to create further synthetics. We discuss three such applications in this section.

Application 1: Creating a Synthetic Bond

Suppose a trader would like to buy a t_1 –maturity bond at time t_0. The trader also wants this bond to be liquid. Unfortunately, he discovers that the only bond that is liquid is an on-the-run Treasury with a longer maturity of t_2. All other bonds are off-the-run. How can the trader create the liquid short-term bond synthetically assuming that all bonds are of discount type and that, contrary to reality, forward loans are liquid?

Rearranging Equation (17), we get

Long t_1-maturity bond	=	Forward loan from t_1 to t_2	+	Short $B(t_0,t_1)$ / $B(t_0, t_2)$ units of t_2 – maturity bond	(19)

The minus sign in front of a contract implies that we need to reverse the position. Doing this, we see that a t_1 – maturity bond can be constructed synthetically by arranging a forward loan from t_1 to t_2 and then by going long $\frac{B(t_0,t_1)}{B(t_0,t_2)}$ units of the bond with maturity t_2. The resulting cash flows would be identical to those of a short bond. More importantly, if the forward loan and the long bond are liquid, then the synthetic will be more liquid than any existing off-the-run bonds with maturity t_1. This construction is shown in figure 4-4.

Application 2:Covering a Mismatch

Consider a bank that has a maturity mismatch at time t_0. The bank has borrowed t_1-maturity funds from Euromarkets and lent them at maturity t_2. Clearly, the bank has to roll over the short-term loan that becomes due at time t_1 with a new loan covering the period $[t_1, t_2]$. This new loan carries an (unknown) interest rate L_{t1} and creates a mismatch risk. The contractual equation in Formula (18) can be used to determine a hedge for this mismatch, by creating a synthetic forward loan, an, in this fashion, locking in time – t_1 funding costs.

In fact, we know from the contractual equation in Formula (18) that there is a relationship between short and long maturity loans:

t_2-9maturity loan	=	Forward loan from t_1 to t_2	+	t_2 – maturity deposit.	(20)

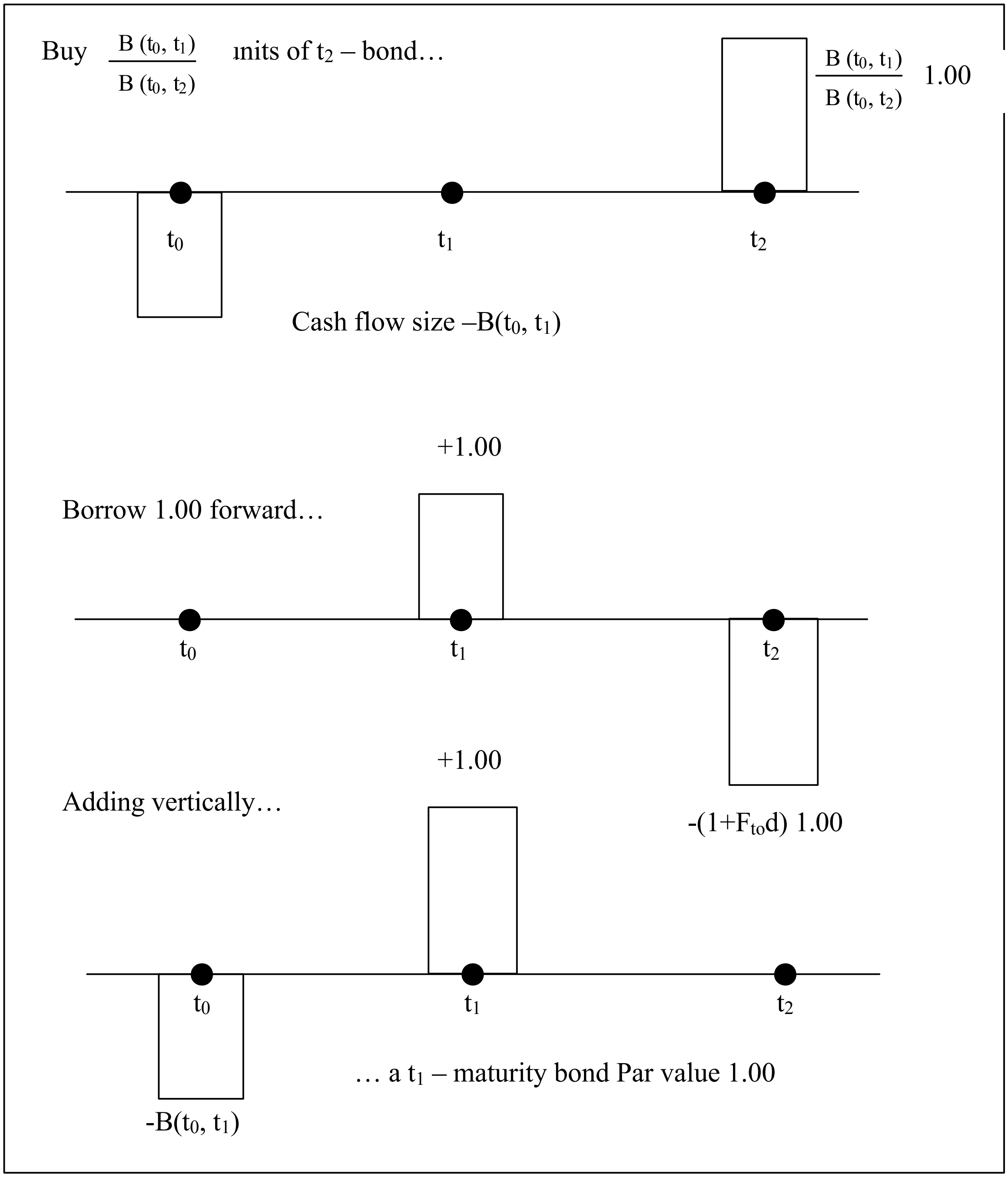
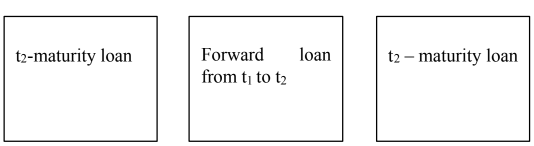


FIGURE 4

Changing signs, this becomes



(21)

According to this the forward loan converts the short loan into a longer maturity loan and this way eliminates the mismatch.

Application 3: Yield Curve Strategies

Suppose we believe that the spread between short rates and long rates will narrow (widen). In other words, you anticipate a flattening (steepening) of the yield curve. What kind of strategy can benefit from these expected shifts?

A reasonable strategy is to take a position that receives the current long rate and that pays the current short rate. As the yield curve flattens, the long rate will decline relative to the short rate and the position will benefit from this movement. This means that the practitioner should borrow short-term and lend long term. It is important to realize that the "bet" is not on the absolute level of interest rates. In fact, yield curve can become flatter, by shifting up or down. The bet is on relative changes. A contractual equation can be used to construct this position efficiently.

Consider the following bond – market contractual equation:

Forward loan from t_1 to t_2	=	Short a proper amount of t_2 – maturity bonds	+	Buy t_1 – maturity bond

(22)

This equation illustrates an interesting point. The yield curve flattening strategy mentioned earlier involved paying the short rate and receiving the long rate. The right-hand side of the contractual equation in (22) does exactly this, except for the signs. Our trader needs to pay short rates and receive long rates. Changing signs we get

Forward loan from t_1 to t_2	=	Buy a proper amount of t_2 – maturity bonds	+	Short a proper amount of t_1 – maturity bond

(23)

By shorting a t_1 maturity bond one pays “short” rates, and by buying a t_2 maturity bond one receives “long” rates. According to the contractual equation this is equivalent to a forward deposit from t_1 to t_2. thus instead of taking the positions directly using bonds, the practitioner can take the same yield curve flattening position using forward rates. All he or she needs to do is lend forward, or even better, just lock in the forward rate using proper instruments. This latter strategy will avoid almost all of the undesirable characteristics of using bonds or deposits during the replication process. After all, cash bonds will have serious balance sheet effects. The following section shows how this approach can be put together very efficiently using forward rate agreements.

Forward Rate Agreements

A forward loan contract implies not one but two obligations. First, 100 units of currency will have to be received at time t_1, and second, interest F_{t0} has to be paid. One can see several drawbacks to such a contract:

1. The forward borrower may not necessarily want to receive cash at time t_1. in most hedging and arbitraging activities, the players are trying to lock in an unknown interest rate and are not necessarily in need of “cash”. A case in point is the convergence play described in section 2, where practitioners were receiving (future) Italian rates and paying (future) Spanish rates. In these strategies, the objective of the players was to take a position on Spanish and Italian interest rates. None of the parties involved had any wish to end up with a loan in one or two years.
2. A second drawback is that forward loan contracts involve credit risk. It is not a good idea to put a credit risk on balance sheet if one wanted to lock in an interest rate.
3. These attributes may make speculators and arbitrageurs stay away from any potential forward loan markets, and the contract may be illiquid.

These drawbacks make the forward loan contract a less-than-perfect financial engineering instrument would separate the credit risk and the interest rate commitment that coexist in the forward loan. It turns out that there is a nice way this can be done.

Eliminating the Credit Risk

First, note that a player using the forward loan only as a tool to lock in the future Libor rate L_{t1} will immediately have to relent the USD100 received at time t_1 at the going market rate L_{t1}. Figure 4-5a displays a forward loan committed at time t_0. Figure 4-5b shows the corresponding spot deposit. The practitioner waits until time t_1 and then makes a deposit at the rate L_{t1}, which will be known at that time. This way, the practitioner cancels an obligation to receive 100 and ends up with only the fixed rate F_{t0} commitment.

Thus, the joint use of a forward loan, and a spot deposit to be made in the future, is sufficient to reach the desired objective – namely, to eliminate the risk associated with the unknown Libor rate L_{t1}. These steps will lock in F_{t0}. we consider the result of this strategy in Figure 4-5c. Add vertically the cash flows of the forward loan (4-5a) and the spot loan (4-5b). Time-t_1 cash flows cancel out since they are in the same currency. Time-t_2 payment and receipt of the principal will also cancel. What are left the respective interest payments. This means that the portfolio consisting of

{A forward loan for t_1 initiated at t_0, a spot deposit at t_1} (24)

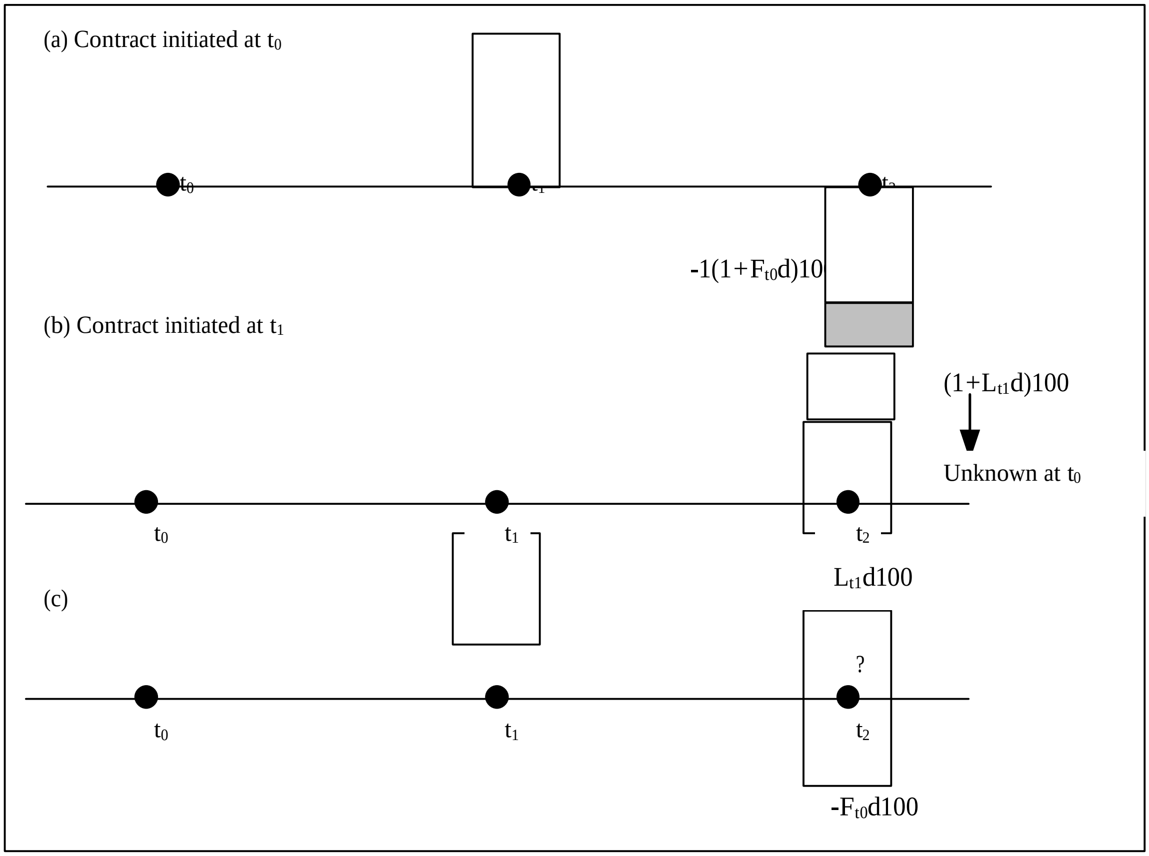
Will lead, according to Figure 5c, to the following (net) cash flows:

	Cash paid	Cash received	Total
Time t_1	-100	+100	0
Time t_2	$-100(1+F_{t0}d)$	$100(1+L_{t1}d)$	$100(L_{t1} - F_{t0})d$

Thus, letting the principal of the forward loan be denoted by the parameter N, we see that the portfolio in expression (24) results in a time $-t_2$ net cash flow equaling

$$N\,(L_{t1} - F_{t0})\,d \qquad (25)$$

Where d is the day’s adjustment to interest, as usual.



Fuigure 5

Definition of the FRA

This is exactly where the FRA contract comes in. If a client has the objective of locking in the future borrowing or lending costs using the portfolio in (24), why not offer this to him or her in a single contract? This contract will involve only the exchange of two interest payments shown in Figures 5c.

In other words, we write a contract that specifies notional amount, N, the dates t_1 and t_2, the "price" F_{t0}, with payoff N (L_{t1} – F_{t0}) d . This instrument is a paid-in-arrears forward rate agreement or a FRA. In a FRA contract, the purchaser accepts the receipt of the following sum at time t_2:

$$(L_{t1} - F_{t0})\, dN \tag{26}$$

if $L_{t1} > F_{t0}$ at date t_1. On the other hand, the purchaser pays

$$(F_{t0} - L_{t1})\, dN \tag{27}$$

if $L_{t1} > F_{t0}$ at date t_1. Thus, the buyer of the FRA will pay fixed and receive floating.

In the case of market-traded FRA contracts, there is one additional complication. The settlement is not done in-arrears at time t_2. Instead, FRAs are settled at time t_1, and the transaction will involve the following discounted cash flows. The

$$\frac{(L_{t1} - F_{t0})\, dN}{1+L_{t1}d} \tag{28}$$

will be received at time t_1, if $L_{t1}>F_{t0}$ at date t_1. On the other hand.

$$\frac{(F_{t0} - L_{t1})\, dN}{1+L_{t1}d} \tag{29}$$

will be paid at time t1, if $L_{t1} < F_{t0}$. Settling at t_1 instead of t_2 has one subtle advantage for FRA seller, which is often a bank. If during [t_0, t_1] the interest rate has moved in favor of the bank, time-$_{t2}$ settlement will reduce the marginal credit risk associated with the payoff. The bank can then operate with a lower credit line.

An Interpretation

Note one important interpretation. A FRA contract can be visualized as an exchange of two interest payments. The purchaser of the FRA will be paying the known interest $F_{t0}\,d$ N and is accepting the (unknown) amount $L_{t0}\,d$ N. Depending on which one is greater, the settlement will be a receipt or a payment. The sum $F_{t0}\,d$ N can be considered, as of time t_0, as the fair payment market participants are willing to make against the random and unknown $L_{t0}\,d$ N. It can be regarded as the "market value" of $L_{t0}\,d$ N.

FRA Contractual Equation

We can immediately obtain a synthetic FRA using the ideas displayed in Figure 5. Figure 5 displays a swap of a fixed rate loan of size N, against a floating rate loan of the same size. Thus, we can write the contractual equation

Buying a FRA	=	Fixed rate loan starting t_1 ending t_2	+	Floating rate deposit starting t_1 ending t_2	(30)

It is clear from the construction in Figure 5, that the FRA contract eliminates the credit risk associated with the principals – since the two N's will cancel out – but leaves behind the exchange of interest rate risk. In fact, we can push this construction further by "plugging in" the contractual equation for the fixed rate forward loan obtained in Formula (18) and get.

Buying a FRA	=	Fixed rate loan starting t_1 ending t_2	+	Floating rate deposit starting t_1 ending t_2	+	Spot deposit starting t_1 ending t_2	(31)

This contractual equation can then be exploited to create new synthetics. One example is the use of FRA strips.

Application: FRA Strips

Practitioner's use portfolios of FRA contracts to form FRA strips. These in turn can be used to construct synthetic loans and deposits and help to hedge swap positions. The best way to understand FRA strips is with an example that is based on the contractual equation for FRAs obtained earlier.

Suppose a market practitioner wants to replicate a 9-month fixed-rate borrowing synthetically. Then the preceding contractual equation implies that the practitioner should take a cash loan at time t_0, pay the Libor rate L_{t0}, and buy a FRA strip made of two sequential FRA contracts, a (3X6)FRA and a(6x9)FRA. This will give a synthetic 9-month fixed – rate loan. Here the symbol (3x6) means t_2 is 6-months and t_1 is 3-months.

Futures: Eurocurrency Contracts

Forward loans do not trade in the OTC market because FRAs are much more cost-effective. Eurocurrency futures are another attractive alternative. In this section, we discuss Eurocurrency futures using the Eurodollar futures as an example and then compare it with FRA contracts. This comparison illustrates some interesting aspects of successful contract design in finance.

FRA contracts involve exchanges of interest payments associated with a floating and a fixed rate loan. The Eurodollar futures contracts trade future loans indirectly. The settlement will be in cash and the futures contract will again result only in an exchange of interest rate payments. However, there are some differences with the FRA contracts.

Eurocurrency futures trade the forward loans (deposits) shown in Figure 1 as homogenized contracts. These contracts deal with loans and deposits in Euromarkets, as suggested by their name. The buyer of the Eurodollar futures contract is a potential depositor of 3-month Eurodollars and will lock in a future deposit rate.

Eurocurrency futures contracts do not deliver the deposit itself. At expiration date t_1, the contract is cash settled. Suppose we denote the price of the futures contract quoted in the market by Q_{t0}. then the buyer of a 3-month Eurodollar contract "promises" to deposit $100(1-\tilde{F}_{t0}1/4)$ dollars at expiration date t_1 and receive 100 in 3 months. The implied annual interest rate on this loan is then calculated by the formula.

$$= \tilde{F} \quad \frac{100.00 - Q_{t0}}{100.00 - Q_{t0}} \tag{32}$$

This means that the price quotations are related to forward rates through the formula.

$$Q_{t0} = 100.00(1- \quad F_{t0}) \tag{33}$$

However, there are important differences with forward loans. The interest rate convention used for forward loans is equivalent to a money market yield. For example, to calculate the time – t_1 present value at time t_0 we let

$$PV(t_0, t_1, t_2) = \frac{100}{1+F_{t0}d} \tag{34}$$

Futures contracts, on the other hand, use a convention similar to discount rates to calculate the time –t_1 values of the forward loan.

$$\hat{PV}(t_0, t_1, t_2) = 100(1 - d\tilde{F}_{t0}) \tag{35}$$

If we want the amount traded to be the same:

$$PV(t_0, t_1, t_2) = PV(t_0, t_1, \hat{t}_2) \tag{36}$$

The two forward rates on the right-hand side of Formulas (34) and (35) cannot be identical. Of course, there are many reasons for the right-hand side and left-hand side in Formula (36) not to be the same. Futures markets have mark-to-markets; FRA markets, in general, do not. With mark-to-market, gains and losses occur daily, and these daily cash flows may be correlated with the overnight funding rate. Thus, the forward rates obtained from FRA markets need to be adjusted to get the forward rate in the Eurodollar futures, and vice versa.

Example

Suppose at time t_0, futures markets quote a price

$$Q_{t0} = 94.67 \tag{38}$$

For a Eurodollar contract that expires on third Wednesday of December 2002. this would two mean two things. First, the implied forward rate for that period is given by:

$$F_{t0} = \frac{100.00 - 94.67}{100} = 0.0533 \tag{39}$$

Second, the contract involves a position on the delivery of

$$100(1-0.0533/4) = 98.67 \tag{40}$$

dollars on the third Wednesday of December 2002.

At expiry these funds will never be deposited explicitly, instead, the contract will be cash settled. For example, if on expiration the exchange has set the delivery settlement price at $Q_{t1} = 95.60$, this would imply a forward rate

$$F_{t0} = 100.00 - 94.67 \quad = \frac{0.0533}{100}$$

and a settlement

$$100(1- 0.0440/4) = 98.90 \tag{42}$$

Thus, the buyer of the original contract will be compensated as if he or she is making a deposit of 98.67 and receiving a loan of 98.90. The net gain is

$$98.90 - 98.67 = 0.23 \quad \text{per 100 } \textit{dollars} \tag{43}$$

This gain can be explained as follows. When the original position was taken, the (forward) rate for the future 3-month deposit was 5.33%. Then at settlement this rate declined to 4.4%

Actually, the above example is a simplification of reality as the gains would never be received as a lump-sum at the expiry due to marking-to-market. The mark-to-market adjustments would lead to a gradual accumulation of this sum in the buyer's account. The gains will earn some interest as well. This creates another complication. Mark-to-market gains losses may be correlated with daily interest rate movements applied to these gains (losses).

Other Parameters

There are some other important parameters of futures contracts. Instead of discussing these in detail, we prefer to report contract descriptions directly. The following table describes this for the CME Eurodollar contract.

Delivery months	: March, June, September, December (for 10 years)
Delivery (Expiry) Day	: Third Wednesday of delivery month
Last trading day	: 11.00 Two business days before expiration
Minimum tick	: 0.0025 (for spot-month contract)
"Tick value"	: USD 6.35
Settlement rule	: BBA Libor on the settlement date

The design and the conventions adopted in the Eurodollar contract may seem a bit odd to the reader, but the contract is a successful one. First of all, quoting Q_{t0} instead of the forward rate $\tilde{F}_{t0}$ makes the contract similar to buying and selling a futures contract on T-bills. This simplifies related hedging and arbitrage strategies. Second, as mentioned earlier, the contract is settled in cash. This way, the functions of securing a loan and locking in an interest rate are successfully separated.

Third, the convention of using a linear formula to represent the relationship between Q_{t0} and $\tilde{F}_{t0}$ is also a point to note. Suppose the underlying time-t_1 deposit is defined by the following equation.

$$D(t_0, t_1, t_2) = 100(1- \quad d) \quad \tilde{F}_{t0} \tag{44}$$

A small variation of the forward rate $\tilde{F}_{t0}$ will result in a constant variation in $D(t_0, t_1, t_2)$:

$$\frac{\partial D(t_0, t_1, t_2)}{\partial \tilde{F}_{t0}} = -8100 = -25 \tag{45}$$

Thus, the sensitivity of the position with respect to the underlying interest rate risk is constant, and the product is truly linear with respect to $\tilde{F}_{t0}$.

The "TED Spread"

The difference between the interest rates on Treasury Notes (T-Notes) and Eurodollar (ED) futures is called the TED spread. T-Note rates provide a measure of the U.S. government's medium term borrowing costs. Eurodollar futures relate to short-term private sector borrowing costs. Thus the "TED spread" has credit risk elements in it. Traders form strips of Eurodollar futures and trade them against T-Notes of similar maturity. A similar spread can be put together using Treasury Bills (T-Bills) and Eurodollars as well.

Given the different ways of quoting yields, calculation of the spread involves some technical adjustments. T-Notes use bond equivalent yields whereas Eurodollars are quoted similar to discount rate basis. The calculation of the TED spread requires putting together strips of futures while adjusting for these differences. There are several technical points that arise along the way.

Once the TED spread is calculated, traders put on trades to benefit from changes in the yield curve slope and in private sector credit risk. For example, traders would long the TED spread if they expected the yield spread to widen. In the opposite case, they would short the TED spread and would thus benefit from the narrowing of the yield spread.

Comparing FRAs and Eurodollar Futures

A brief comparison of FRAs with Eurocurrency futures may be useful. (1) Being OTC contracts, FRAs are more flexible instruments, since Eurodollar futures trade in terms of preset homogeneous contracts. (2) FRAs have the advantage of confidentially. There is no requirement that the FRA terms be announced. The terms of a Eurocurrency contract are known. (3) There are, in general, no margin requirements for FRAs and the mark-to-market requirements are less strict. With FRAs, money changes hands only at the settlement date. Eurocurrency futures come with margin requirements as well as with mark-to-market requirements are less strict. With FRAs, money changes hands only at the settlement date. Eurocurrency futures come with margin requirements as well as with mark-to-market requirements. (4) FRAs have counterparty risk, whereas the credit risk of Eurocurrency futures contracts are insignificant. (5) FRAs are quoted on an interest rate basis while Eurodollar futures are quoted on a price basis. Thus a trader who sells a FRA will hedge this position by selling a Eurodollar contract as well. (6) Finally, an interesting difference occurs with respect to fungibility. Eurocurrency contracts are fungible, in the sense that contracts with the same expiration can be netted against each other even if they are entered into at different times and for different purposes. FRA contracts cannot be netted against each other even with respect to the same counterparty, unless the two sides have a specific agreement.

Convexity Differences

Besides these structural differences, FRAs and Eurocurrency futures have different convexities. The pricing equation for Eurocurrency futures is linear in $\tilde{F}_{t0}$, whereas the market traded FRAs have a pricing equation that is nonlinear in the corresponding Libor rate. We will see that this requires convexity adjustments, which is one reason why we used different symbols to denote the two forward rates.

Hedging FRAs with Eurocurrency Futures

For short-dated contracts, convexity and other differences may be negligible, and we may ask the following question. Putting convexity differences aside, can we hedge a FRA position with futures, and vice versa?

It is best to answer this question using an example. The example also illustrates some real world complications associated with this hedge.

Example

Suppose we are given the following Eurodollar futures prices on June 17, 2002: September price (delivery date: September 16) 96.500 (implied rate = 3.500)

December price (delivery date: December 16) 96.250 (implied rate = 3.750)

March price (delivery date: March 17) 96.000 (implied rate = 4.000)

A trader would like to sell a (3 x 6) FRA on June 17, with a notional amount of USD 100,000,000. How can the deal be hedged using these futures contracts?

Note first that according to the value and settlement date conventions, the FRA will run for the period September 19 through December 19 and will encompass 92 days. It will settle against the Libor fixed on September 17. the September futures contract, on the other hand, will settle against the Libor fixed on September 16 and is quoted on a 30/360 basis. Thus, the implied forward rates will not be identical for this reason as well.

Let f be the FRA rate and ª be the differnces between this rate and the forward rate implied byu the futures contract. Using formula (28), the FRA settlement, with notional value of 100 million USD, may be written as

$$\frac{100m\,((0.035 + ?\,) - \text{Libor})\,92/360}{(1+ \text{Libor}\,92/360)} \tag{46}$$

Note that this settlement is discounted to September 19 and will be received once the relevant Libor rate becomes known. Ignoring mark-to-market and other effects, a futures contract covering similar periods will settle at

$$a\ (am\ (0.0350 - \text{Libor})\ 90/360) \tag{47}$$

Note at least two differences. First, the contract has a nominal value of USD 1million. Second, 1 month is, by convention, taken as 30 days, while in the case of FRA it was his actual number of days. The á is the number of contracts that has to be chosen so that the FRA positoni is correctly hedged.

The trader has to choose á such that the two settlement amounts are as close as possible. This way, by taking opposite positions in these contracts, the trader will hedge his or her risks.

Some Technical Points

The process of hedging is an approximation that may face several technical and practical difficulties. To illustrate them we look at the preceding example once again.

1. Suppose we tried to hedge (or price) a strip of FRAs rather than a single FRA be adjusted to contract using a strip of available futures contracts. Then the strip of FRAs will have to deal with increasing notional amounts. Given that futures contracts have fixed notional amounts, contract numbers need to be adjusted instead.
2. As indicated, a 3-month period in futures markets is 90days, whereas FRA contracts count the actual number of days in the corresponding 3-month period.
3. Given the convexity differences in the pricing formulas, the forward rates implied by the two contracts are not eh same and, depending on Libor volatility, the ª may be large or small.
4. There may be differences of 1 or 2 days in the fixing of the Libor rates in the two contracts.

There technical differences relate to this particular example, but they are indicative of most hedging and pricing activity.

Real-World Complications

Up to this point, the discussion ignored some real-life complications. We made the following simplifications. (1) We ignored bid-ask spreads. (2) Credit risk was assumed away. (3) We ignored the fact that the fixing date in an FRA is, in general, different from the settlement date. In fact this is another date involved in the FRA contract. Let us now discuss these issues.

Bid-Ask Spreads

We being with bid-ask spreads. The issue will be illustrated using a bond market construction. When we replicate a forward loan via the bond market, we buy a B (t_0, t_1) bond and short-sell a B (t_0, t_2) bond. Thus, we have to use ask prices for B (t_0, t_1) and bid prices for B (t_0,t_2). This means that the asking price for a forward interest rate will be

$$1+F_{10}^{ask}d = \frac{B(t_0,t_1)^{ask}}{B(t_0,t_2)^{bid}} \tag{48}$$

Similarly, when the client sells a FRA, he or she has to use the bid price of the dealers and brokers. Again, going through the bond markets we can get

$$1+F_{10}^{ask}d = \frac{B(t_0,t_2)^{bid}}{B(t_0,t_1)^{ask}} \tag{49}$$

This means that

$$F_{t0}^{bid} < F_{t0}^{ask} \tag{50}$$

The same bid-ask spread can also be created from the money market synthetic using the bid-ask spreads in the money markets.

$$1 + F_{t0}^{ask}\,d = \frac{1+ L_{t0}^{1ask}\,d^1}{1+ L_{t0}^{2ask}\,d^2} \tag{51}$$

Clearly, we again have

$$F_{t0}^{bid} < F_{t0}^{ask} \quad (52)$$

Thus, pricing will normally yield two-way prices.

In market practice, FRA bid-ask spreads are not obtained in the manner shown here. The bid-ask quotes on the FRA rate are calculated by first obtaining a rate from the corresponding Libors and then adding a spread to both sides of it. Many practitioners also use the more liquid Eurocurrency futures to "make" markets.

An Asymmetry

There is another aspect to using FRAs for hedging purposes. The net return and net cost from an interest rate position will be asymmetric since, whether you buy (pay fixed) or sell (receive fixed), a FRA always settles against Libor. But Libor is an offer (asking) rate, and this introduces an asymmetry.

We begin with a hedging of floating borrowing costs. When a company hedges a floating borrowing cost, both interest rates from the cash and the hedge will be Libor based. This means that:

1. The company pays Libor + margin to the bank that it borrows funds from.
2. The company pays the fixed FRA rate to the FRA counterparty for hedging this floating cost.
3. Against which the company receives Libor from the FRA counterparty.

Adding all receipts and payments, the net borrowing cost becomes *FRA rate + margin*.

Now consider what happens when a company hedges, say, a 3-month floating receipt. The relevant rate for the cash position is Libid, the bid rate placing funds with the Euromarkets. But a FRA always settles always against Libor. So the picture will change to

- Company receives Libid, assuming a zero margin.
- Company receives FRA rate.
- Company pays Libor.

Thus, the net return to the company will become FRA – (Libor-Libid).

Forward Rates and Term Structure.

A detailed framework for fixed income engineering will be discussed in Chapter 15. However, some preliminary modeling of the term structure is in order. This will clarify the notation and some of the essential concepts.

Bond Prices

Let $\{B(t_0, t_1), I = 1,2……,n\}$ represent the bond price family, where each $B(t_0,t_i)$ is the price of a default – free zero – coupon bond that matures and pays $1 at time t_i. These $\{B(t_0, t_i)\}$ can also be viewed as a vector of discounts that can be used to value default – free cash flows.

For example, given a complicated default-free asset, A_{t0}, which pays deterministic cash flows $\{C_{ti}\}$ occurring at arbitrary times, t_i, i= 1,….., k, we can obtain the value of the asset easily if we assume the following bond price process:

$$A_{t0} = \sum C_{ti} B)t_0, t_i \quad (53)$$

That is to say, we just multiply the t_i^{th} cash flow with the current value of one unit of currency that belongs to t_i, and then sum over i.

This idea has an immediate application in the pricing of a coupon bond. Given a coupon bond with a nominal value of $1 that pays a coupon rate of c% at times t_i, the value of the bond can easily be obtained using the preceding formula, where the last cash flow will include the principal as well.

What Forward Rates Imply

In this lesson, we obtained the important arbitrage equality

$$1+ F(t_0, t_1, t_2)\, d = \frac{B(t_0, t_1)}{B(t_0, t_2)} \quad (54)$$

where the $F(t_0, t_1, t_2)$ is written in the expanded form to avoid potential confusion. It implies a forward rate that applies to a loan starting at t_1 and ending at t_2. writing this arbitrage relationship for all the bonds in the family $\{B(t_0, t_1),\}$, we see that

$$1+ F(t_0, t_0, t_1)\, d = \frac{B(t_0, t_0)}{B(t_0, t_1)} \quad (55)$$

$$1+ F(t_0, t_1, t_2)\, d = \frac{B(t_0, t_1)}{B(t_0, t_2)} \quad (56)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots \quad (57)$$

$$1+ F(t_0, t_{n-1}, t_n)\, d = \frac{B(t_0, t_{n-1})}{B(t_0, t_n)} \quad (58)$$

Successively substituting the numerator on the right-hand side using the previous equality and noting that for the first bond we have $B(t_0, t_0) = 1$, we obtain

$$B(t_0,t_n) = \frac{1}{(1 + F(t_0, t_0, t_1)\, d) \ldots (1+ F(t_0, t_{n-1}, t_n)\, d)} \quad (59)$$

We have obtained an important result. The bond price family $\{B(t_0, t_i)\}$ can be expressed using the forward rate family,

$$\{F(t_0, t_0, t_1), \ldots\ldots F(t_0, t_{n-1}, t_n)\} \quad (60)$$

Therefore if all bond prices are given we can determine the forward rates.

Remark

Note that the "first" forward rate $F(t_0, t_0, t_1)$ is contracted at time t_0 and applies to a loan that starts at time t_0. Hence, it is also the t_0 spot rate:

$$\frac{1}{B(t_0, t_1)}$$

$$(1 + F(t_0, t_0, t_1)\, d) = (1+L_{t0}\, d) = \quad (61)$$

$$B(t_0, t_1) = \frac{1}{(1+L_{t0}d)} \tag{62}$$

The bond price family $B(t_0, t_i)$ is the relevant discounts factors that market practitioner use in obtaining the present values of default – free cash flows. We see that modeling F_{t0}'s will be quite helpful in describing the modeling of the yield curve or, for that matter, the discount curve.

Conventions

FRAs are quoted as two-way prices in bid-ask format, similar to Euro deposit rates. A typical market contributor will quote a 3-month and a 6-month series.

Example

The 3-month series will look like this:

1×4	4.87	4.91
2×5	4.89	4.94
3×6	4.90	4.95
etc.		

The first row implies that the interest rates are for a 3-month period that will start in 1 month. The second row gives the forward rate for a loan that begins in 2 months for a period of 3 months and so on.

The 6 – month series will look like this:

1×7	4.87	4.91
2×8	4.89	4.94
3×9	4.90	4.95
etc.		

According to this table, if a client would like to lock in a fixed payer rate in 3 months for a period of 6 months and for a notional amount of USD 1 million, he or she would buy the 3s against 9s and pay the 4.95% rate. For 6 months, the actual net payment of the FRA will be

$$\frac{1{,}000{,}000\,((L_{t3}/100) - 0.0495)1/2}{(1 + \tfrac{1}{2}\times L_{t3}/100)} \tag{63}$$

where L_{t3} is the 6-month Libor rate that will be observed in 3 months.

Another convention is the use of Libor rate as a reference rate for both the sellers and the buyers of the FRA. Libor being an asking rate, one might think that a client who sells a FRA may receive rate than Libor. But this is not true, as the reference rate does not change.

A Digression: Strips

Before finishing this chapter we discuss an instrument that is the closest real life equivalent to the default – free pure discount bonds $B(t_0, t_i)$. This instrument is called strips. U.S. strips have been available since 1985 and U.K. strips since 1997.

Consider a long – term straight Treasury bond, a German bund, or a British gilt and suppose there are no implicit options. These bonds make coupon payments during their life at regular intervals. Their day-count and coupon payment intervals are somewhat different, but in essence they are standard long-term debt obligations. In particular, they are not the zero-coupon bonds that we have been discussing in this chapter.

Strips are obtained from coupon bonds. The market practitioner buys a long-term coupon bond and then "strips" each coupon interest payment and the principal and trades them separately. Such bonds will be equivalent to zero – coupon bonds except that, if needed, one can put them back together and reconstruct the original coupon bond.

The institution overseeing the government bond market, the Bank of England in the United Kingdom or the Treasury in the United States, arranges the necessary infrastructure to make stripping possible and also designates the strippable securities. Note that only some particular dealers are usually allowed to strip and to reconstruct the underlying bonds. These dealers put in a request to strip a bond that they already have in their account and then they sell the pieces separately. As an example, a 10-year gilt is strippable into 20 coupons plus the principal. There will be 21 zero-coupon bonds with maturities 6, 12, 18, 24 (and so on) months.

Exercises

1. You have purchased 1 Eurodollar contract at a price of Q_0 = 94.13, with an initial margin of 5%. You keep the contract for 5 days and then sell it by taking the opposite position. In the meantime, you observe the following settlement prices:

   $\{Q_1 = 94.23, Q_2 = 94.03, Q_3 = 93.93, Q_4 = 93.43, Q_5 = 93.53\}$

   a. Calculate the string of mark-to-market losses or gains.

   b. Suppose the spot interest rate during this 5-day period was unchanged at 6.9%. What is the total interest gained or paid on the clearing firm account?

   c. What are the total gains and losses at settlement?

2. The treasurer of a small bank has borrowed funds for 3 months at an interest rate of 6.73% and has lent funds for 6 months at 7.87%. the total amount is USD38 million.

   To cover his exposure created by the mismatch of maturities, the dealer needs to borrow another USD38 million for months, in 3 months' time, and hedge the position now with a FRA.

   The market has the following quotes from three dealers:

BANK A	3×6	6.92-83
BANK B	3×6	6.87-78
BANK C	3×6	6.89-80

   a. What is (are) the exposure(s) of this treasurer? Represent the result on cash flow diagrams.

   b. Calculate this treasurer's break-even forward rate of interest, assuming no other costs.

   c. What is the best FRA rate offered to this treasurer?

   d. Calculate the settlement amount that would be received (paid) by the treasurer if on the settlement date, the Libor fixing was 6.09%.

3. A corporation will receive USD7 million in 3 months' time for a period of 3 months. The current 3-month interest rate quotes are 5.67 to 5.61. The Eurodollar futures price is 94.90.

   Suppose in 3 months the interest rate becomes 5.25% for 3-month Euro deposits and the Eurodollar futures price is 94.56.

   a. How many ticks has the futures price moved?
   b. How many futures contracts should this investor buy or sell if she wants to lock in the current rates?
   c. What is the profit (loss) for an investor who bought the contract at 94.90?

4. Suppose practitioners learn that the British Banker's Association (BBA) will change the panel of banks used to calculate the yen Libor. One or more of the "weaker" banks will be replaced by "stronger" banks at a future date.

   The issue here is not whether yen Libor will go down, as a result of the panel now being "stronger". In fact, due to market movements, even with stronger banks in the panel, the yen Libor may in the end go up significantly. Rather, what is being anticipated is that the yen Libor should decrease in London relative to other yen fixing s, such as Tibor. Thus, to benefit from such a BBA move, the market practitioner must form a position where the risks originating from market movements are eliminated and the "only" relevant variable remains the decision by the BBA.

   a. How would a trader benefit from such a change without taking on too much risk?
   b. Using cash flow diagrams, show how this can be done.
   c. In fact, show which spread FRA position can be taken. Make sure that the position is (mostly) neutral toward market movements and can be created, the only significant variable being the decision by the BBA.

(From IFR, issue 1267) Traders lost money last week following the British Bankers' Association (BBA) decision to remove one Japanese bank net from t eh yen Libor fixing panel. The market had been pricing in no significant changes to the panel just the day before the changes were announced.

Prior to the review, a number of dealers were reported to have been short the Libor/Tibor spread by around 17 bp, through a twos into fives forward rae agreement (FRA) spread contract. This was in essence a bet that the Japanese presence on the Libor fixing panel would be maintained.

When the results of the review were announced on Wednesday January 20, the spread moved out by around 22 bp – leaving the dealers with mark-to-market losses. Some were also caught out by a sharp movement in the one-year yen/dollar Libor basis swap, which moved in from minus 26bp to minus 14 bp.

The problems for the dealers were caused by BBA's decision to alter the nature of the fixing panel, which essentially resulted in one Japanese bank being removed to be replaced by a foreign bank. Bank of China, Citibank, Tokai Bank and Sakura wee taken out, while Deusche Bank, Norinchukin Bank, Rabobank and WestLB were added.

The move immediately increased the overall credit quality of the grouping of banks responsible for the fixing rate. This caused the yen Libor fix – the average cost of panel banks raising funds it he yen money market – to fall by 8bp in a single day. Dealers said that one Japanese bank was equivalent to a 5 bp lower yen Libor rate and that the removal of the Bank of China was equivalent to a 1 bp or 2 bp reduction.

Away from the immediate trading losses, market reaction to the panel change was mixed. The move was welcome by some, who claimed that the previous panel was unrepresentative of the yen cash business being done.

"Most of the cash is traded in London by foreign banks. It doesn't make sense to have half Japanese banks on the panel", said one yen swaps dealer. He added that because of the presence of a number of Japanese banks on the panel, yen Libor rates were being pushed above where most of the active yen cash participants could fund themselves in the market.

Others, however, disagreed. "it's a domestic [Japanese] market at the end of the day. The BBA could now lose credibility in Japan," said one US bank money markets trader.

BBA Officials said the selections were made by the BBA's FX and Money Markets Advisory Panel, following private nominations and discussions with the BBA Libor Steering Group. They said the aim of the advisory panel was to ensure that the contributor panels broadly reflected the "balance of activity in the interbank deposit market".

INTRODUCTION TO SWAP ENGINEERING

Objectives

- After studying this lesson you will be able to understand why we buy and sell securities when you can swap the corresponding return and achieve the same objective efficiently and at minimum cost.

Dear friends, you see, financial institutions, investors, or corporations do not have emotional attachment to physical stocks, bonds, or credits. Instead, their rationale is either to gain exposure to stock prices of interest rates or to hedge risks associated with them. Investors want to receive stock or bond returns, whereas financial companies desire to hedge interest rates, currency, or credit risk. Commodities are, of course, needed as raw materials, but even here the purpose often is not to buy or sell the commodity itself, but to buy and sell the *risk* associated with its price.

If the objective is to have exposure to interest rates or towards equity indices, or to hedge them, why should decision makers buy or sell the physical asset itself, especially when there is a better alternative? Note that buying and selling cash asset s has many negatives. First of all, they require or generate cash. Hence, positions should be funded or the cash generated should be placed. This will affect balance sheets. Second, when assets are bought or sold, the operations may generate capital gains or losses. This has tax implications. Regulations also interfere with the actual purchase and sale of various assets. Third, the asset itself may not be that easy to buy or to sell. For example, to receive the return associated with the S&P 500, an investor needs to buy 500 stocks and then periodically rebalance his portfolio to ensure correct tracking of the index. This may be a costly task. Finally, when some assets are bought outright in order to have exposure to the corresponding returns, the investor surrenders his or her cash, and this may create a credit risk.

Swaps can accomplish the same functions without any of these negatives. Swaps do not require any outright cash payments. Instead of buying a bond, one simply receives the corresponding interest – but pays a floating market rate in return. Because desired exposure is gained by exchanging cash flows, swaps do not have credit risk. Finally, when the desired exposure is gained through swaps, tax and regulatory considerations often change and may become friendlier to the counterparties.

There is another reason why swaps are one of the central concepts in financial engineering. The logic in engineering various types of swaps is becoming a benchmark in its own right. The methodology used in swap markets may very well be the most market friendly way of going about the pricing, hedging, and risk management practices.

1.1. An Example from Banking

We begin with a simple question. What is a bank? The answer depends on the type of bank we are concerned with. Money center banks are complex institutions that have a variety of functions. Many of these involve sophisticated finance, mathematics, and numerical methods. Neighborhood banks, or bank branches, on the other hand, conduct relatively simple services such as issuing letters of credit, credit cards, or mortgages and wiring receiving funds for their customers. These activities may not be as complex as investment banking or trading and, may carry relatively low margins, but they are essential for the daily functioning of the economy. Take loans, for example, Consumer loans for customers who buy cars, and business loans for financing investments, exports, and imports are essential banking activities.

What is the function of such a neighborhood bank that deals with retail customers? These institutions do not take positions on the direction of the stock market, or of interest rates. The expertise of the local bank on these maters hardly extends beyond that of laymen. Neighborhood banks do not make markets in forward rate agreements (FRAs) or foreign-exchange (FX) forwards, or any other complex derivatives for that matter. Instead, a neighborhood bank evaluates households or firms, and then makes loans, say, to finance the purchase of a car, or to provide working capital. A bank that is “good” at this particular activity is one that knows how to evaluate its customers’ credit.

In other words, from a financial engineering perspective traditional banking activity should not take positions on the direction of interest rates. Once this becomes clear, there are two problems. First, retail customers desire loans that have long maturities and, in general, at fixed interest rates. On the other hand, when customers deposit money, they prefer to do this for (relatively) short maturities. Thus, traditional banks may sometimes have to conduct business by agreeing to receive a fixed interest rate during a long period of time while paying an interest that may change in a relatively short period of time while paying an interest that may change in a relatively short period of time. This involves an interest rate risk.

There is another reason why banks may take such risks. In general, yield curves are upward sloping. This means short-term loans carry low rates of interest, whereas long-term funds are more expensive. Thus, if the T-bill rates are, say, at 2%, the longer-term 10 year bond yields may be around, say, 4.5%. A private credit pays more. For example, in lending to another bank, the interbank market may demand an additional spread of 40 basis points. This makes the cost of short-term money equal to

$$\text{Benchmark 6-month rate} + \text{credit spread} = 2.00\% + 40 \text{ bp} = 2.40\% \qquad (1)$$

These funds can then be lent out at rates of over 5% for maturities of 10 years and longer.

Hence, there is a natural incentive to borrow short-term from interbank markets and then on-lend long-term to retail clients – especially if the interest rate environment seems stable and if there is little danger of spikes in short-term rates. A neighborhood bank may not have the necessary research department to forecast future interest rate behavior, yet, when the yield curve is "steep", the bank management may be tempted to borrow short-term and lend long-term. This will also create an interest rate risk.

Interest rate swaps (IRS) have emerged as the main tool for hedging and managing such risks. In fact, swaps have become the largest and one of the most liquid of all instrument classes. The swap curve is regarded by many practitioners as a benchmark for yield curve analysis.

2. The Instrument: Swaps

Imagine any two sequences of cash flows with different characteristics. These cash flows could be generated by any process-a financial instrument, a productive activity, a natural phenomenon. They will also depend on different risk factors. Then, one can, in principle, device a contract where these two cash flow sequences are exchanged. This contract will be called a swap. To design a swap, we use the following principles:

1. A swap is arranged as a pure exchange of cash flows and hence should not require any additional net cash payments at initiation. In other words, the initial value of the swap contract should be zero.
2. The contract specifies a swap spread. This variable is adjusted to make the two counterparties willing to exchange the cash flows.

A generic exchange is shown in figure 5-1. In this figure, the first sequence of cash flows starts t time t_1 and continues periodically at $t_2, t_3, \ldots, t_k$. There are k cash flows of differing sizes denoted by

$$\{C(s_{t0}, x_{t1}), C(s_{t0}, x_{t2}), \ldots, C(s_{t0}, x_{tk})\} \quad (2)$$

These cash flows depend on a vector of market or credit risk factors denoted by x_{ti}. The cash flows depend also on the s_{t0}, a swap spread or an appropriate swap rate. By selecting value of s_{t0}, the initial value of the swap can be made zero.

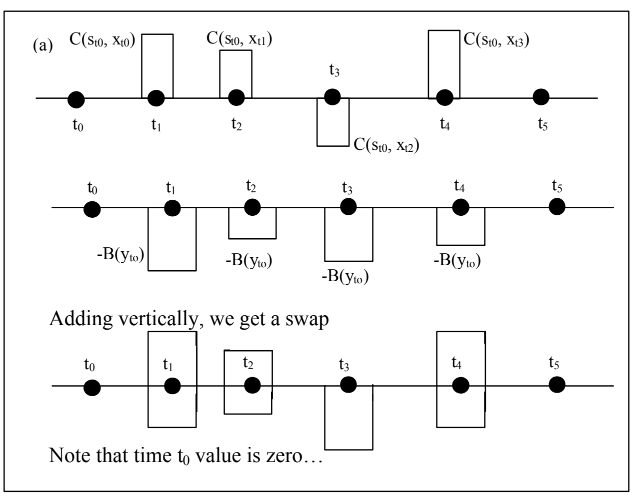


Figure 5.1

Figure 5-1b represents another strip of cash flows:

$$\{B(y_{t0}), B(y_{t1}), B(y_{t2}), \ldots, B(y_{tk})\} \quad (3)$$

which depend potentially on some other risk factors denoted by y_{ti}.

The swap consists of exchanging the $\{C(s_{t0}, x_{ti})\}$ against $\{B(y_{ti})\}$ at settlement dates $\{t_i\}$. The parameter s_{t0} is selected at time t_0 so that the two parties are willing to go through with exchange without any initial cash payment. This is shown in figure 5-1c. One will pay the C (.)'s and receive the B(.)'s. The counterparty will be the "other side" of the deal and will do the reverse. Clearly, if the cash flows are in the same currency, there is no need to make two different payments in each period t_i. One party can simply pay the other the net amount. Then actual wire transfers will look more like the cash flows in Figure 5-2. Of course, what one party receives is equal to what the counterparty pays.

Now, if two parties who are willing to exchange the two sequences of cash flows without any up-front payment, the market value of these cash flows must be the same no matter how different they are in terms of implicit risks. Otherwise one of the parties will require an up-front net payment. Yet, as time passes, a swap agreement may end up having a positive or negative net value, since the variable x_{ti} and y_{ti} will change, and this will make one cash flow more "valuable" than the other.

Example

Suppose you singed a swap contract that entitles you to a 7% return in dollars, in return for a 6% return in Euros. The exchanges will be made every 3 months at a predetermined exchange rate e_{t0}. at initiation time t_0, the net value of he commitment should be zero, given the correct swap spread.

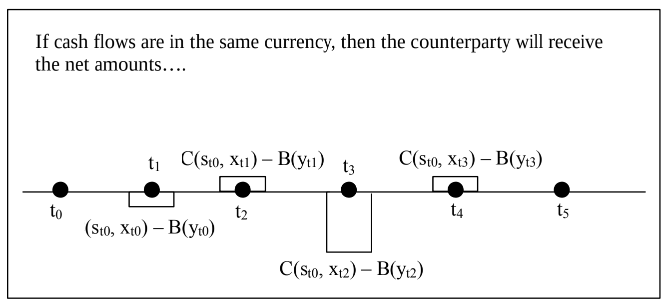


Figure 5.2

This means that at time t_0 the market value of he receipts and payments are the same. Yet, after the contract is initiated, USD interest rates may fall relative to European rates. This would make the receipt of 7% USD funds relatively more valuable than the payments in Euro.

As a result, from the point of view of the USD – receiving party, the value of the swap will move from zero to positive, while for the counterparty the swap will have a negative value.

Of course, actual exchanges of cash flows at times $t_1, t_2, \ldots, t_n$ may be a more complicated process than the simple transactions shown in Figure 5-2. What is exactly is paid or received? Based on which price? Observed when? What are the penalties if deliveries are not made on time? What happens if a t_i falls on a holiday? A typical swap contract needs to clarify many such parameters. These and other issues are specified in the documentation set by the international swaps and derivatives associations.

1. Types of Swaps

Swaps are a very board instrument category. Practically, every cash flow sequence can be used to generate a swap. It is impossible to discuss all the relevant material in this book. So, instead of spreading the discussion thinly, we adopt a strategy where a number of critical swap structures are selected and the discussion is centered on these. We hope that the extension of the implied swap engineering to other swap categories will be straightforward.

3.1. Non – Interest Rate Swaps

Most swaps are interest rate related given the Libor and yield curve exposures on corporate and bank balance sheets. But swaps form a broader category of instruments, and to emphasize this point we start the discussion with non-interest rate swaps. Here the most recent and the most important is the Credit default Swap. We will examine this credit instrument in a separate chapter, and only introduce it briefly here. This chapter will concentrate mainly on two other swap categories: equity swaps and commodity swaps.

3.1.1. Equity Swaps

One swap category involves the exchange of returns from an equity index against the return from another asset, often against Libor-based cash flows. These are called equity swaps.

In equity swaps, the parties will exchange two sequences of cash flows. One of the cash flow sequences will be generated by dividend and capital gains (losses), while the other will depend on a money market instrument, in general Libor. Once clearly defined, each cash flow can be valued separately. Then, adding or subtracting a spread to the corresponding Libor rate would make the two parties willing to exchange these cash flows with no initial payment. The contract that makes this exchange legally binding is called an equity swap.

Thus, a typical equity swap consists of the following. Initiation time will be t_0 .An equity index It_i and a money market rate, say Libor Lt_i, are selected. At times $\{t_1, t_2, \ldots\ldots, t_n\}$ the parties will exchange cash flows based on the percentage change in It_i , written as

$$Nt_{i-1} \frac{It_i - It_{i-1}}{I_{ti-1}} \qquad (4)$$

Against Libor-based cash flows, $Nt_{i-1} Lt_{i-1}\, d$ plus or minus a spread. The Nt_i is the notional amount, which is not exchanged.

Note that the notional amount is allowed to be reset at every t_o, $t_1, \ldots\ldots, t_{n-1}$,

Parties to adjust their position in the particular equity index periodically. In equity swaps, this notional principal can be selected as a constant, N.

Example

In Figure 5-3 we have a 4-year sequence of capital gains (losses) plus dividends generated by a certain equity index. They are exchanged every 90 days, against a sequence of cash flows based on 3-month Libor-20bp.

The notional principal is USDI million. At time t_o the element of these cash flows will be unknown.

At time t_1 the respective payments can be calculated once the index performance is observed. Suppose we have the following data:

$$It_o = 800 \qquad (5)$$

$$It_1 = 850 \quad Lt_o = 5\% \; spread = .20 \qquad (6)$$

Then the time-t_1 equity-linked cash flow is

$$1m \frac{It_1 - It_0}{It_0} = 1{,}000{,}000(0.0625) = 62{,}500 \qquad (7)$$

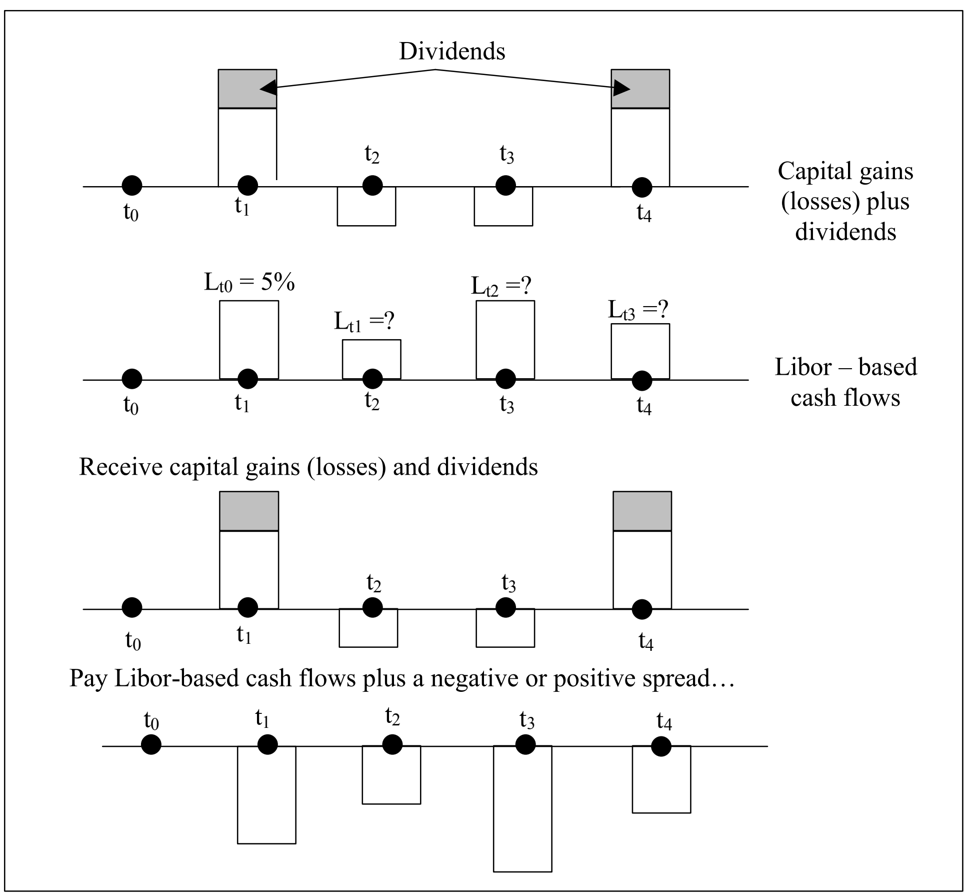


FIGURE 5-3

The Libor-linked cash flows will be

$$1m\,(L_{t_0} - s_{t_0})\frac{90}{360} = 1{,}000{,}000(.05 - .002)\frac{1}{4} = 12{,}000 \quad (8)$$

The remaining unknown cash flows will become known as time passes, dividends are paid prices move. The spread is subtracted from the interest rate.

Some equity swaps are between two equity indices. The following example illustrates the idea.

Example

In an equity swap, the holder of the instrument pays the total return of the S&P 500 and receives the return on another index, say the Nikkei. Its advantage for the holder lies in the fact that, as a swap, it does not involve paying any up-front premium.

Of course, the same trade could also be created by selling S&P future and buying futures on another equity index. But, the equity swap has the benefit that it simplifies tracking the indices.

Later in this lesson, we will discuss several uses of equity swaps.

3.1.1. Commodity Swaps

The overall structure of commodity swaps is similar to equity swaps. As with equity swaps, there are two major types of commodity swaps. Parties to the swap can, either (1) exchange fixed to floating payments based on a commodity index or, (2) exchange payments when one payment is based on an index and the other on a money market rate.

Consider a refinery, for example. Refineries buy crude oil and sell refined products. They may find it useful to lock in a fixed price for crude oil. This way, they can plan future operations better. Hence, using a swap, a refiner may want to receive a floating price of oil and pay a fixed price per barrel.

Such commodity or oil swaps can be arranged for all sorts of commodities, metals, precious metals, and energy prices.

Example

Japanese oil companies and trading houses are naturally short in crude oil and long in oil products. They use the short-term swap market to cover this exposure and to speculate, through the use of floating/fixed-priced swaps. Due to an over-capacity of heavy oil refineries in the country, the Japanese are long in heavy-oil products and short in light-oil products. This has produced a swap market of Singapore light-oil products against Japanese heavy-oil products.

There is also a "paper balance" market, which is mainly based in Singapore but developing in Tokyo. This is an oil instrument, which is settled in cash rather than through physical delivery of oil. (IFR, Issue 946)

The idea is similar to equity swaps, so we prefer to delay further discussion of commodity swaps until we present the exercises at the end of this chapter.

3.1.2. Credit Swaps

This is an important class of swaps, and it is getting more important by the day. There are many variants of credit swaps, and they will be discussed in more detail in a separate chapter. Here we briefly introduce the main idea, which follows the same principle as other swaps. The credit default swap is the main tool for swapping credit. We discuss it briefly in this chapter.

If swaps are exchanges of cash flows that have different characteristics, then we can consider two sequences of cash flows that are tied to two different credits.

A 4-year floating rate cash flow made of Libor plus a credit spread is shown in Figure 5-4. The principal is USD 1 million and it generates a random cash flow. But there is a critical difference here relative to the previous examples. Since the company may default, there is no guarantee that the interest or the principal will be paid back at future dates. Figure 5-4a simplifies this by assuming that the only possible default on the principal is at time t_4, and that simplifies this by assuming that the only possible default on the principal is at time t_4, and that when the default occurs all the principal and

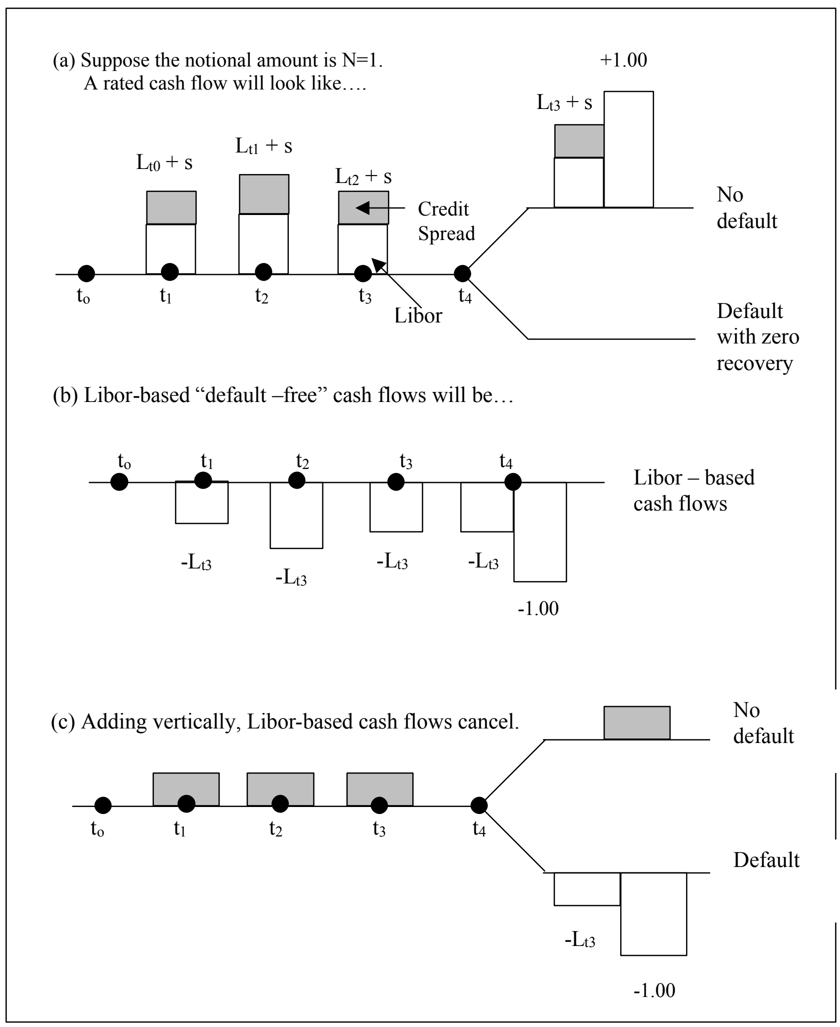


Figure 5.4

interest is lost. Figure 5-4b displays a default – free market cash flow based on 6 month Libor.

By adding the two sequences of cash flows in this example vertically, we get the credit default swap, shown in Figure 5-4.

3.1. Interest Rate Swaps

This is the largest swap markets. It involves exchanging cash flows generated by different interest rates. The most common case is when a *fixed* swap rate is paid (received) against a floating Libor rate in the *same* currency. Interest rate swap have become a fundamental instrument in world financial markets. The following reading illustrates this for the case of plain vanilla interest rate swaps.

Example

The swap curve is being widely touted as the best alternative to a dwindling to a dwindling Treasury market for benchmarking U.S. corporate bonds... This has prompted renewed predictions that the swap curve will be adopted as a primary benchmark for corporate bonds and asset-backed securities.

... Investors in corporate bonds say there are definite benefits from the increasing attention being paid to swap spreads for valuing bonds. One is that the mortgage-backed securities market has already to a large degree made the shift to use of Libor-based valuation of positions, and that comparability of corporate bonds with mortgage holdings is desirable.

... Swaps dealers also point out that while the agency debt market is being adroitly positioned by Fannie Mae and Freddie Mac as an alternative to the Treasury market for benchmarking purposes, agency spreads are still effectively bound to move in line with swap spreads.

... Bankers and investors agree that hedging of corporate bond positions in the future will effectively mean making the best use of whatever tools are available. So even if swaps and agency bonds have limitations, and credit costs edge up, they will still be increasingly widely used for hedging purposes. (IFR, Issue 6321)

This reading illustrates the crucial position held by the swap market in the world of finance. The “swap curve” obtained from interest rate swaps is considered by many as a benchmark for the term structure of interest rates, and this means that most assets could eventually be priced off the interest rate swap, in one way or another. Also, the reading correctly points out some major sectors in markets. In particular, (1) the mortgage-backed securities (MBS) market, (2) the market for “agencies”, which means securities issued by Fannie Mae or Freddie Mac, etc., and (3) the corporate bond market have their own complications, yet, swaps play a major role in all of them. At this point, it is best to define formally the interest rate swap and then look at an example.

A plain vanilla interest rate swap (IRS) initiated at time t_0 is a commitment to exchange interest payments associated with a notional amount N, settled at clearly identified settlement dates, $\{t_1, t_2, \ldots, t_n\}$. The buyer of the swap will make fixed payments of size s_{t0}Nä each and receive floating payments of size L_{ti}Nä. The Libor rate L_t is determined at set dates $\{t_0, t_1, \ldots, t_{n-1}\}$. The maturity of swap is m years. The s_{t0} is the swap rate.

Example

An Interest Rate Swap has a notional amount N of 1 million USD, a 7% fixed rate for 2 years in semiannual (s.a.) payments against a cash flow generated by 6-monoth Libor rates will be observed during this period. The L_{t0} is known at the initial point t_0. The remaining Libor rates, L_{t1}, L_{t2}, and L_{t3}, will be observed gradually as time passes but are unknown initially.

In Figure 5-5, the floating cash flows depending on L_t, are observed at time t_i, but are paid-in-arrears at times t_{i+1}. Swaps that have this characteristic are known as paid-in-arrears swaps.

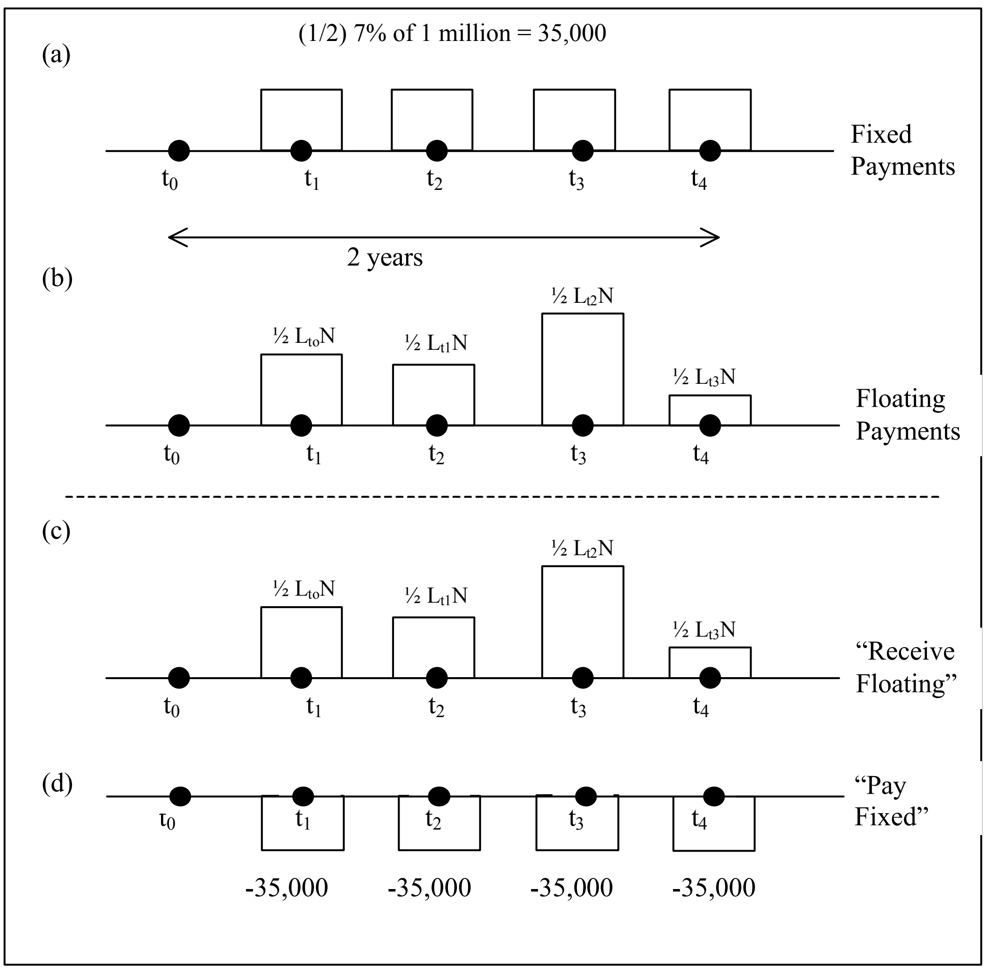


FIGURE 5-5

Clearly, we have two sets of cash flows with different market risk characteristics. The market will price them separately. Once this is done, market participants can trade them. A *fixed payer* will pay the cash flows in Figure 5-5a and receive the one in Figure 5-5b. This institution is the buyer of the interest rate swap.

The market participant on the other side of the deal will be doing the reverse – receiving cash flows based on a fixed interest rate at time t_0, while paying cash flows that become gradually known as time passes and the Libor rates L_{ti} are revealed. This party is the fixed receiver, whom the market also calls the seller of the swap. We can always make the exchange of the two cash flows acceptable to both parties by adding a proper spread to one of the cash flows. This role is played by the swap spread. The market includes the spread in the fixed rate. By adjusting this spread accordingly, the two parties may be brought together and accept the exchange of one cash flow against another. The agreed fixed rate is the swap rate. We have:

$$\text{Swap rate} = \text{Benchmark rate} + \text{Swap spread} \qquad (9)$$

The *benchmark rate* is often selected as the same maturity sovereign bond in that currency.

The final cash flows of an interest rate swap from the fixed payer's point of view will be as shown in Figure 5-2. Only the net amount will change hands.

A real-life example might be helpful. In the following, we consider a private company that is contemplating an increase in the proportion of its floating rate debt. The company can do this by issuing short – term paper, called commercial paper (CP), and continuing to roll over the debt when these obligations mature. But a second way of doing it is by first issuing a 5-year fixed-rate bond, and then swapping the interest paid into floating interest rates.

Example

A corporation considers issuing commercial paper or a medium - term fixed – rate bond (MTN) that it can convert to a floating – rate liability via a swap. The company is looking to increase the share of floating – rate liabilities to 50% -55% from 30%.

The alternative to tapping the MTN market is drawing on its $700 million commercial paper facility.

This reading shows one role played by swaps in daily decisions faced by corporate treasuries. The existence of swaps makes the rates observed in the important CP-sector more closely related to the interest rates in the MTN-sector.

3.1.1. Currency Swaps

Currency swaps are similar to interest rate swaps, but there are some differences. First, the exchanged cash flows are in different currencies. This means that two different yield curves are involved in swap pricing instead of just one. Second, in the large majority of cases a floating rate is exchanged against another floating rate. A third difference lies in the exchange of principals at initiation and a re-exchange at maturity. In the case of interest rate swaps this question does not arise since the notional amounts are in the same currency. Currency swaps can be engineered almost the same way as interest rate swaps.

Formally, a currency swap will have the following components. There will be two currencies, say USD ($) and Euro (ª). The swap is initiated at time t_0 and involves (1) an exchange of a principal amount $N^{\$}$ against the principal m^{a} and (2) a series of floating interest payments associated with the principals $N^{\$}$ and M^{a}, respectively. They are settled at settlement dates, $\{t_1, t_2, \ldots, t_n\}$. One party will pay the floating payments $L^{\$}_{ti} N^{\$}$ä and recieve floating payments of size $L^{a}_{ti} M^{a}$ä. The two Libor rates $L^{\$}_{ti}$ and L^{a}_{ti} will be determined at set dates $\{t_0, t_1, \ldots, t_{n-1}\}$. The maturity of swap will be *m* years.

A small *spread* s_{t0} can be added to one of the interest rates to make both parties willing to exchange the cash flows. The market maker will quote bid/ask rates for this spread.

Example

Figure 5-6 shows a currency swap. The USD notional amount is 1 million. The current USD/EUR exchange rate is at 0.95. The agreed spread is 6 bp. The initial 3-month Libor rates are

$$L^{\$}_{ti} = 3\% \qquad (10)$$

$$L^{a}_{ti} = 3.5\% \qquad (11)$$

This means that at the first settlement date

$$(1{,}000{,}000)\,(0.03 + 0.0006)\,¼ = \$7{,}650 \qquad (12)$$

will be exchanged against

$$0.95\,(1{,}000{,}000)\,(0.035)\,¼ = {}^{a}8312.5 \qquad (13)$$

All other interest payments would be unknown. Note that the Euro principal amount is related to the USD principal amount according to

$$e_{t0} N^{\$} = M^{a} \qquad (14)$$

where e_{t0} is the spot exchange rate at t_0.

Also, note that we added the swap spread to the USD Libor.

Pricing currency swaps will follow the same principles as interest rate swaps. A currency swap involves well-defined cash flows and consequently we can calculate an arbitrage-free value for each sequence of cash flows. Then these cash flows are traded. An appropriate spread is added to either floating rate.

By adjusting this spread, a swap dealer can again make the two parties willing to exchange the two cash flows.

3.2.2 Basis Swaps

Basis swaps are similar to currency swaps except that often there is only one currency involved. A basis swap involves exchanging cash flows in one *floating* rate, against cash flows in another *floating* rate, in the same currency. One of the involved interest rates is often a non-Libor-based rate, and the other is Libor.

The following reading gives an idea about the basis swap. Fannie Mae, a U.S. government agency, borrows from international money markets in USD Libor and then lends these funds to mortgage banks. Fannie Mae faces a *basis risk* while doing this. There is a small difference between the interest rate that it eventually pays, which is USD Libor, and the interest rate it eventually receives, the USD discount rate. To hedge its position, Fannie Mae needs to convert one floating rate to the other. This is the topic of the reading that follows:

Example

Merrill Lynch has been using Fannie Mae benchmark bonds to price and hedge its billion dollar discount/basis swap business. "We have used the benchmark bonds as a pricing tool for our discount/Libor basis swaps since the day they were issued. We continue to use them to price the swaps and hedge our exposure," said the head of interest rate derivatives trading.

He added the hedging activity was centered on the five and 10-year bonds – the typical discount/Libor basis swap tenors. Discount/Libor swaps and notes are employed extensively by U.S. agencies, such as Fannie Mae, to hedge their basis risk. They lend at the U.S. discount rate but fund themselves at the Libor rate and as a result are exposed to the Libor / discount rate spread. Under the basis swap, the agency / municipality receive Libor and in return pays the discount rate.

Major U.S. derivative providers began offering discount / Libor basis swaps several years ago and now run billion dollar books. (IFR, Issue 1229)

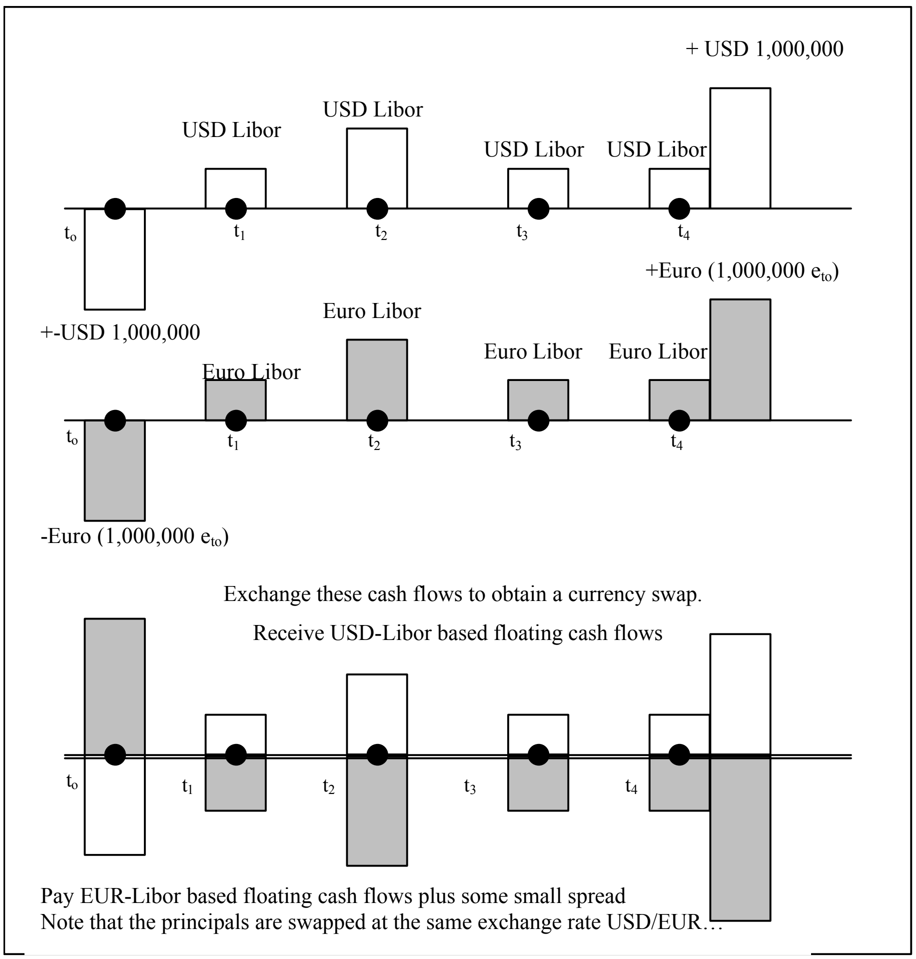


FIGURE 5-6

This reading illustrates two things. Fannie Mae needs to swap one floating rate to another in order to allow the receipts and payments to be based on the same risk. But at the same time, *because* Fannie Mae is hedging using basis swaps and because there is a large amount of such Fannie Mae bonds, some market practitioners may think that these agency bonds make good pricing tools for basis swaps themselves.

3.1.1. What is an Asset Swap?

The term *asset swap* can, in principle, be used for any type of swap. After all, sequences of cash flows considered thus far are generated by some assets, indices, or references rates. Also, swaps linked to equity indices or reference rates such as Libor can easily be visualized a s floating Rate Notes (FRN), corporate bond portfolios, or portfolios of stocks. Exchanging these cash flows is equivalent to exchanging the underlying asset.

Yet, the term asset swap is often used with a more precise meaning. Consider a default able *par* bond that pays annual coupon $c_{to.}$ Suppose the payments are semi-annual. Then we can imagine a swap where coupon payments are exchanged against 6-month Libor L_{ti} plus a spread $s_{to,}$ every 6-months. The coupon payment are fixed and known at t_o. *The floating payments will be random, although the spread component,* $s_{to,}$ is known at well. This structure is often labeled an asset swap. The reader can easily put together the cash flows implied by this instrument, if the issue of default is ignored. Such a cash flow diagram would follow the exchanges shown in Figure 5-1. One sequence of cash flows would represent coupons, the other Libor plus a spread.

Asset swaps interpreted this way offer a useful alternative to investors. An investor can always buy a bond and receive the coupon $c_{to.}$ But by using an asset swap, the investor can also swap out of the coupon payments and receive only floating Libor Plus the spread $s_{to.}$ This way the exposure to the *issuer* is kept and the exposure to *fixed interest rates* is eliminated. In fact, treasury bonds or fixed receiver interest swaps may be better choices if one desires exposure to fixed interest rates. Given the use of Libor in this structure, the s_{to} is calculated as the spread to the corresponding fixed swap rates.

3.2.2 More Complex Swaps

The swaps discussed thus far are liquid and are traded actively. One can imagine many other swaps. Some of these are also liquid, others are not. Amortizing swaps, bullion swaps, MBS swaps, quanto (differential) swaps are some the come to mind. We will not elaborate on them at this point; some of these swaps will be introduced as examples or exercises in later chapters.

An interesting special case is constant maturity swaps (CMS), which will be discussed in detail in Chapter 15. The CMS swaps have an interesting convexity dimension that requires taking into account volatilities and correlations across various forward rates along a yield curve. A related swap category is constant maturity treasury (CMT) swaps.

3.2. Swap Conventions

Interest rate swap markets have their own conventions. In some economies, the market quotes the swap *spread*. This is the case of USD interest rate swaps. USD interest rates swaps are quoted as a spread to Treasuries. In Australia, the market also quotes swap spreads. But the spreads are to bond futures.

In other economies, the market quotes the swap rate. This is the case for Euro interest rate swaps.

Next, there is the issue of how to quote swaps. This is done in terms of two-way interest rate quotes. But sometimes the quoted swap rate is on an annual basis, and sometimes it is on a semiannual basis. Also, the day-count conventions change from one market to another. In USD swaps, the day-count is in general ACT/360. In EUR swaps day-count is 30/360.

According to market conventions, a fixed payer called the payer, is long the swap, and has bought a swap. On the other hand, a fixed receiver called the *receiver,* is short the swap, and has sold a swap.

4. Engineering Interest Rate Swaps

We now study the financial engineering of swaps. We focus on plain vanilla *interest rate* swaps. Engineering of other swaps is similar in many ways, and is left to the reader. For simplicity, we deal with a case of only three settlement dates. Figure 5-7 shows a fixed –payer, three – period interest rate swap, with start date t_0. The swap is initiated at date t_0. The party will make three fixed payments and receive three floating payments at date t_1, t, and t_3 are the settlement dates, and the t_0, t_1 and t_2 are the *reset* dates. The latter are dates on which the relevant Libor rate is determined.

We select the notional amount N as unity and let ä=1, assuming that the floating rate is 12 – month Libor.

$$N = 1\$ \qquad (15)$$

Under these simplified conditions the fixed payments equal st_0, and the Libor-linked payments equal L_{t0}, L_{t1}, and L_{t2}, respectively. The *swap spread* will be the difference between s_{t0} and the treasury rate on the bond with the same maturity, denoted by y_{t0}. Thus, we have

$$\text{Swap spread} = s_{t0} - y_{t0} \qquad (16)$$

We will study the engineering of this interest rate swap. More precisely, we will discuss the way we can replicate this swap. More than the exact synthetic, what is of interest is the way(s) one can approach this problem.

A swap can be reverse – engineered in at least two ways:

1. We can first decompose the swap *horizontally,* into two streams of cash flows, one representing a floating stream of payments (receipts), the other a fixed stream. If this is done, then each stream can be interpreted as being linked to a certain type of bond.
2. Second, we can decompose the swap *vertically,* slicing it into n cash exchanges during n time periods. If this is done, then each cash exchange can be interpreted similar (but not identical) to a FRA paid-in-arrears, with the property that the fixed rate is constant across various settlement dates.

We now study each method in detail.

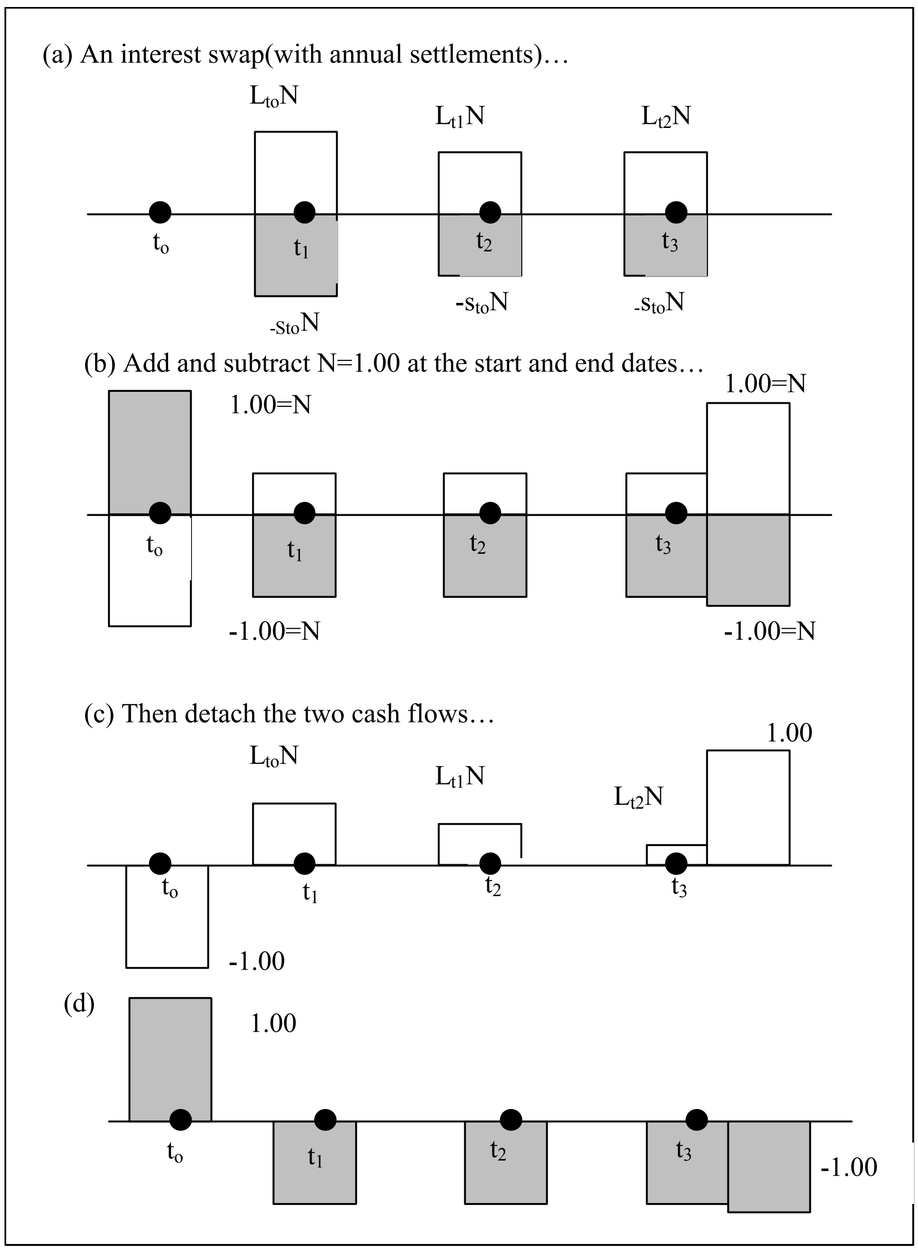


FIGURE 5-7

4.1 A Horizontal Decomposition

First we simplify the notation and the parameters used in this section. To concentrate on the engineering aspects only, we prefer to eliminate some variables from the discussion. For example, we assume that the swap will make payments every year so that the day-count parameter is $d = 1$, unless assumed otherwise. Next, we discuss a *forward* swap that is signed at time t_0, but starts at time t_1, with $t_0 < t_1$. During this discussion, we may occasionally omit the use of the term "forward" and refer to the forward swap simply as a swap.

The traditional way to decompose an interest rate swap is to do this horizontally. The original swap cash flows are shown in Figure 5-7a. Before we start, we need to use a trick. We add and subtract the same notional amount N at the start, and end dates, for both sequences of cash flows. Since these involve identical currencies and identical amounts, they cancel out and we recover the standard exchanges of floating versus fixed-rate payments. With the addition and subtraction of the initial principals, the swap will look as in Figure 5-7 b.

Next, "detach" the cash flows in Figure 5-7b horizontally, so as to obtain two separate cash flows as shown in Figures 5-7c and 5-7d. note that each sequence of cash flows is already in the form of a meaningful financial contract.

In fact, Figure 5-7c can immediately be recognized as representing a long forward position in a floating rate note that pays Libor flat at time t_1, the initial amount N is paid and L_{t1} is set. At t_2, the first interest payment is received, and this will continue until time t_4 where the last interest is received along with the principal.

Figure 5-7d can be recognized as a short forward position on a par coupon bond that pays a coupon equal to s_{t0}. We (short) sell the bond to receive N. At every payment date the fixed coupon is paid and then, at t_4, we pay the last coupon and the principal N.

Thus, the immediate decomposition suggests the following synthetic:

Interest rate swap = {Long FRN with Libor coupon, short par coupon bond} (17)

Here the bond in question needs to have the same credit risk as in a flat Libor-based loan. Using this representation, it is straightforward to write the contractual equation:

Long Swap	=	Short a par bond with Coupon s_{t0}	+	Long FRN paying Libor flat

(18)

Using this representation, it is straightforward to write the contractual equation:

4.1.1. A Synthetic coupon Bond

Suppose an AAA – rated entity with negligible default risk issues only 3-year FRNs that pay Libor-10 bp every 12 months. A client would like to buy a coupon bond from this entity, but it turns out that no such bonds are issued. We can help our client by synthetically creating the bond. To do this, we manipulate the contractual equation so that we have a long coupon bond on the right-hand side:

Long par bond with coupon s_{t0} – 10 bp	=	Sell a swap with rate s_{t0}	+	Long FRN paying Libor- 10bp (19)

(19)

The geometry of this engineering is shown in figure 5-8. The synthetic results in a coupon bond issued by the same entity and paying a coupon of s_{t0} – 10bp. The 10bp included in the coupon account for the fact that the security is issued by an AAA – rated entity.

4.1.2. Pricing

The Contractual equation obtained in (18) permits pricing the swap off the debt markets, using observed prices of fixed and floating coupon bonds. To see this we write the present value of the cash streams generated by the fixed and floating rate bonds using appropriate discount factors. Throughout this section, we will work with a special case of a 3 – year spot swap that makes fixed payments against 12 – month Libor. This simplifies the discussion. It is also straightforward to generalize the ensuing formulas to an n-year swap.

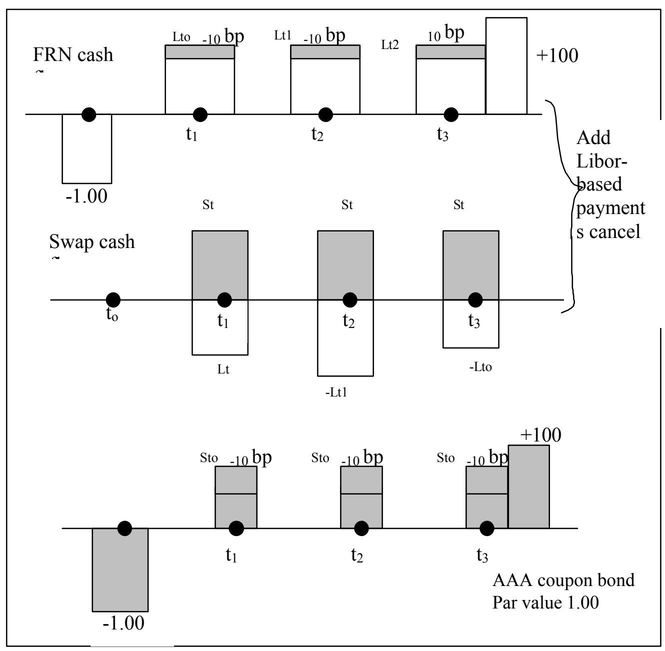


FIGURE 5-8

Suppose the swap makes three annual coupon payments, each equaling $s_{t0}N$. We also have three floating rate payments each with the value $L_{t\,i-1}N$, where the relevant Libor $L_{t\,i-1}$ is set at t_{i-1}, but is paid at t_i.

4.1.3. Valuing Fixed Cash Flows

To obtain the present value of the fixed cash flows, we discount them by the relevant floating rates. Note that, if we knew the floating rates $\{L_{t0}, L_{t1}, L_{t2}\}$, we could write

$$\text{PV-fixed} = \frac{s_{t0}N}{(1+L_{t0})} + \frac{s_{t0}N}{(1+L_{t0})(1+L_{t1})} + \frac{s_{t0}N + N}{(1+L_{t0})(1+L_{t1})(1+L_{t2})} \quad (20)$$

Note that without any loss of generality we added N to date t_3 cash flows. But at t = 0, the Libor rates L_{ti}, i = 1, 2 are unknown. Yet, we know that against each L_{ti}, t=1, 2, the market is willing to pay the known for ward, (FRA) rate, F (t_0, t_i). Thus, using the FRA rates as if they are the time – t_0 market values of the unknown Libor rates, we get

$$\text{PV-fixed as of } t_0 = \frac{s_{t0}N}{(1+F(t_0, t_0))} + \frac{s_{t0}N}{(1+F(t_0, t_0))\,(1+F(t_0, t_1))} + \quad (21)$$

$$\frac{s_{t0}N + N}{(1+F(t_0, t_0))\,(1+F(t_0, t_1))\,(1+F(t_0, t_2))} \quad (22)$$

All the right-hand side quantities are known, and the present value can be calculated exactly, given the s_{t0}.

4.1.4. Valuing Floating Cash Flows

For the floating rate cash flows we have

$$\text{PV-floating as of } t_0 = \frac{L_{t0}N}{(1+L_{t0})} + \frac{L_{t1}N}{(1+L_{t0})(1+L_{t1})} + \frac{L_{t2}N + N}{(1+L_{t0})(1+L_{t1})(1+L_{t2})} \quad (23)$$

Here, to get a numerical answer, we don't even need to use the forward rates. This present value can be written in a much simpler fashion, once we realize the following transformation:

$$\frac{L_{t2}N + N}{(1+L_{t0})(1+L_{t1})(1+L_{t2})} = \frac{(1+L_{t2})N}{(1+L_{t0})(1+L_{t1})(1+L_{t2})}$$

$$\frac{N}{(1+L_{t0})(1+L_{t1})}$$

Then, add this to the second term on the right-hand side of the present value in (23) and use the same simplification,

$$\frac{L_{t1}N}{(1+L_{t0})(1+L_{t1})} + \frac{N}{(1+L_{t0})(1+L_{t1})} = \frac{N}{(1+L_{t0})} \quad (26)$$

Finally, apply the same trick to the first term on the right-hand side of (23) and obtain, somewhat surprisingly, the expression.

$$\text{PV of floating payments as of } t_0 = N \quad (27)$$

According to this, the present value of a FRN equals the par value N at *every settlement date.* Such recursive simplifications can be applied to present values of floating rate payments at reset dates. We can now combine these by letting

$$PV \textit{ of fixed payments as of } t_0 = PV \textit{ of floating payments as of } t_0 \quad (29)$$

This gives an equation where s_{t0} can be considered as an unknown:

$$\frac{s_{t0}N}{(1+F(t_0, t_0))} + \frac{s_{t0}N}{(1+F(t_0, t_0))\,(1+F(t_0, t_1))} + \quad (30)$$

$$\frac{s_{t0}N + N}{(1+F(t_0, t_0))\,(1+F(t_0, t_1))\,(1+F(t_0, t_2))}$$

Canceling the N and rearranging, we can obtain the numerical value of s_{t0} given F (t_0, t_1), and F (t_0, t_2). This would value the swap off the FRA markets.

4.1.5 An Important Remark

Note a very convenient, but *very delicate* operation that was used in the preceding derivation. Using the liquid FRA markets, we "replaced" the unknown L_{ti} by the known $F(t_0, t_i)$ in the appropriate formulas. Yet, these formulas were nonlinear in L_t, and even if the forward rate is an unbiased forecast of the appropriate Libor,

$$F(t_0, t_i) = E^{P^*}_{t0}[L_{ti}] \quad (31)$$

Under some appropriate probability P*, it is not clear whether the substitution is justifiable. For example, it is known that the conditional expectation. Operator at time t_0, represented by $E^{P^*}_{t0}$ [.], cannot be moved *inside* a nonlinear formula due to Jensen's inequality:

$$E^{P^*}_{t0}\left[\frac{1}{1 + L_{ti}d}\right] > \frac{1}{1 + E^{P^*}_{t0}[L_{ti}]\, d} \quad (32)$$

So, it is not clear how L_{ti} can be replaced by the corresponding $F(t_0, t_i)$, even when the relation in (31) is true. These questions will have to be discussed after the introduction of risk-neutral and forward measures in Chapters 11 and 12. Such "substitution" is delicate and depends on many conditions. In our case we are allowed to make the substitution, because the forward rate is what market pays for the corresponding Libor rates at time t_0.

3.1. A vertical Decomposition

We now study the second way of decomposing the swap. We already know what FRA contracts are. Consider an annual FRA where the d = 1. Also, let the FRA be paid-in-arrear. Then, at some time $t_i + d$, the FRA parties will exchange the cash flow:

$$(L_{ti} - F(t_0, t_i))\, d\, N, \quad (33)$$

where N is the notional amount, d = 1, and the F (t_0, t_i) is the FRA rate determined at time t_0. We also know that the FRA amounts to exchanging the fixed payment F (t_0, t_i) d N against the floating payment L_{ti} d N.

Is it possible to decompose a swap into n FRAs, each with a FRA rate $F(t_0, t_i)$, i =1, …,n, ? The situation is shown in Figure 5-9 for the case n =3. The swap cash flows are split by slicing the swap vertically at each payment date. Figures -9b, 5-9c, and 5-9d represent each swap cash flow separately. A fixed payment is made against an unknown floating Libor rate, in each case.

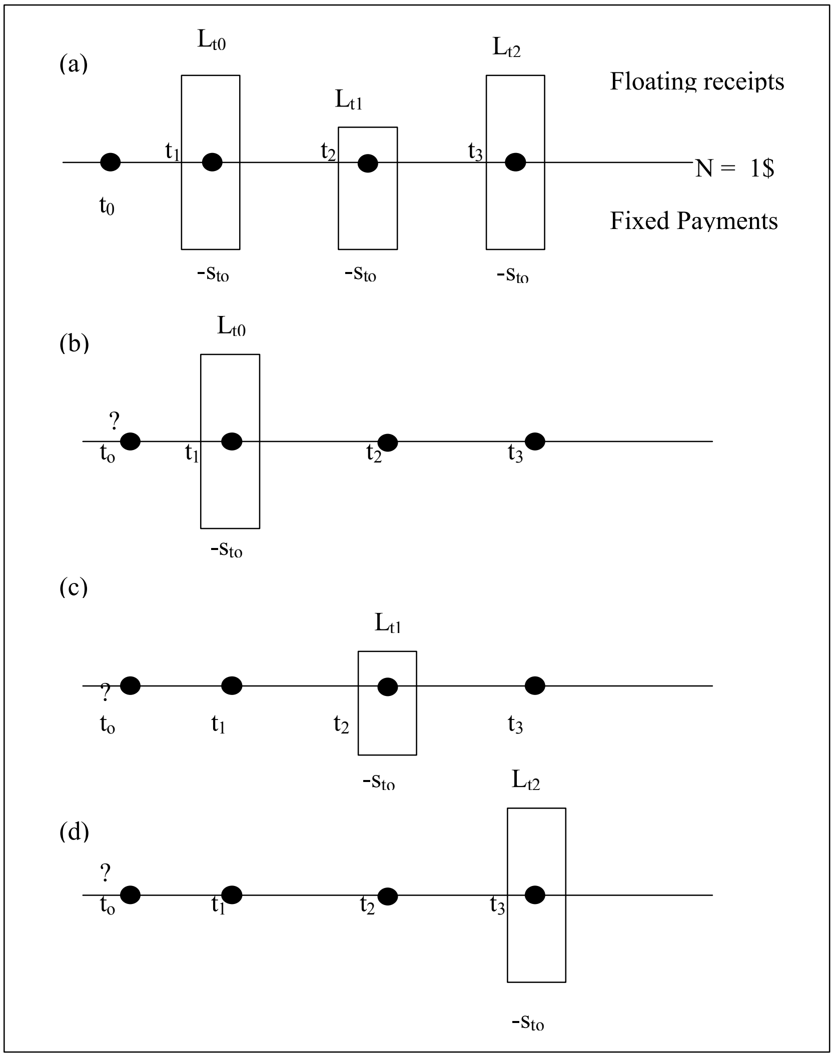


FIGURE 5-9

Are the cash exchanges shown in Figures 5-9b, 5-9c and 5-9d tradeable contracts? In particular, are they valid FRA contracts, so that the swap becomes a portfolio of three FRAs? At first glance, the cash flows indeed look like FRAs. But when we analyze this claim more closely, we see that these cash flows are not valid contracts individually.

To understand this, consider the time – t_2 settlement in Figure 5-9b together with the FRA cash flows for the same settlement date, as shown in Figure 5-10. This figure displays an important phenomenon concerning cash flow analysis. Consider the FRA cash flows initiated at time t_0 and settled in arrears at time t_2 and compare these with the corresponding swap settlement. The two cash flows look similar. A fixed rate is exchanged the same against Libor rate l_{t1} observed at time t_1. But there is still an important difference.

First of all, note that the FRA rate $F(t_0, t_l)$ is determined by supply and demand or by pricing through money markets. Thus, in general

$$F(t_0, t_l) \neq s_{to} \qquad (34)$$

This means that if we buy the cash flow in Figure 5-10a, and sell the cash flow in Figure 5-10b, Libor-based cash flows will cancel out, but the fixed payments won't. As a result, the portfolio will have a known negative or positive net dash flow at time t_2, as shown in Figure 5-10c. Since this cash flow is known exactly, the present value of this portfolio cannot be zero. But the present value of the FRA cash flow is zero, since (newly initiated) FRA contracts do not have any initial cash payments. All these imply that the time t_2 cash flow shown in Figure 5-10c will have a known present value.

$$B(t_0, t_2)[F(t_o, t_1) - s_{to}]\ d\ N, \qquad (35)$$

Where the $B(t_o, t_2)$ is the time t_0 value of the default-free zero coupon bond that matures at time t, with par value USD 1. This present value will be positive or negative depending on whether $F(t_0, t_1) < s_{to}$ or not.

Hence, slicing the swap contract vertically into separate FRA-like cash exchanges does not result in tradeable financial contracts. In fact, the time t_2 exchange shown in Figure 5-10c has a missing time t_0 cash flow of size $B(t_o, t_1)\ [F(t_o, t_1) - s_{to}]$ äN. Only by adding this, does the exchange become a tradeable contract. The s_{to} is a fair exchange for the risks associated with L_{to}, L_{t1}, and L_{t2} "*on the average*". As a result, the time t_2 cash exchange that is part of the swap contract ceases to have a zero present value when considered individually.

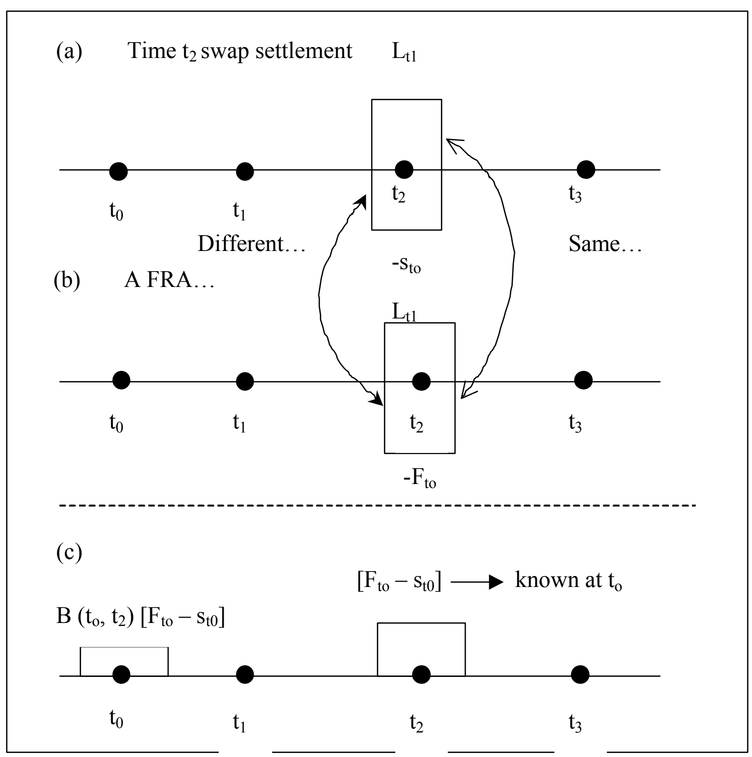


FIGURE 5-10.

4.2.1. Pricing

We have seen that it is not possible to interpret the individual swap settlements as FRA contracts directly. The two exchanges have a nonzero present value, while the (newly initiated) FRA contracts have a price of zero. But the previous analysis is still useful for pricing the swap since it gives us an important relationship.

In fact, we just showed that the time t_2 element of the swap has the present value $B(t_o, t_2)\,[F(t_o, t_1) - s_{to}]\, d\, N$, which is not, normally, zero. This must be true for all swap settlements individually. Yet, the swap cash flows altogether do have zero present value. This leads to the following important pricing relation:

$$B(t_o, t_1)\,[F(t_o,t_o) - s_{to}]\, d\, N + B(t_o, t_2)\,[F(t_o,t_1) - s_{to}]\, d\, N + B(t_o, t_3)\,[F(t_o,t_2) - s_{to}]\, d\, N = 0 \qquad (36)$$

Rearranging provides a formula that ties the swap rate to FRA rates:

$$s_{to} = \frac{B(t_o, t_1)\, F(t_o,t_o) + B(t_o, t_2)\, F(t_o,t_1) + B(t_o, t_3)\, F(t_o,t_2)}{B(t_o, t_1) + B(t_o, t_2) + B(t_o, t_3)} \qquad (37)$$

This means that we can price swaps off the FRA market as well. The general formula, where n is the maturity of the swap,

$$S_{t0} = \frac{\sum^{n}{}_{i=1} B(t_0, t_i) F(t_0, t_{i=1})}{\sum^{n}{}_{i=1} B(t_0, t_i)} \qquad (38)$$

Will be used routinely in this book. It is an important arbitrage relationship between swap rates and FRA rates.

3. Uses of Swaps: Introduction

The general idea involving the use of swaps in financial engineering is easy to summarize. A swap involves exchanges of cash flows. But cash flows are generated by assets, liabilities, and other commitments. This means that swaps are simply a standardized, liquid, and cost-effective alternative to trading cash assets. Instead of trading the cash-asset or liability, we are simply trading the cash flows generated by it. Because swaps, in general, have zero value at the time of initiation, and are very liquid, this will indeed be a cost – effective alternative – hence their use in position taking, hedging, and risk management. What are these uses of swaps? We begin the discussion by looking at equity swaps. We will see that these swaps have convenient balance sheet implications, as seen for the FX-swap in chapter 3. Regulatory and tax treatment of equity swaps are also relevant.

3.1. Uses of Equity Swaps

Equity swaps illustrate the versatility of swap instruments.

3.1.1. Fund Management

There is a huge industry of fund management where the fund manager tries to track some equity index. One way to do this is by buying the underlying portfolio of stocks that replicates the index and constantly readjusting it as the market moves, or as new funds are received, or paid by the fund. This is a fairly complex operation. Of course, one can use the S & P 500 futures to accomplish this. But futures contracts need to be rolled over and they require mark-to market adjustments. Using equity swaps is a cost –effective alternative.

The fund manager could enter into an equity swap using the S & P 500 in which the fund will pay, quarterly, a Libor-related rate and a (positive or negative) spread and receive the return on the S & P 500 index for a period of n years.

The example below is similar to the one seen earlier in this chapter. Investors were looking for cost – effective ways to diversify their portfolios.

Example

In one equity swap, the holder of the instrument pays the total return on the S&P 500 and receives the return on the FRSE 100. Its advantage for the investor is the fact that, as a swap, it does not involve paying any up-front premium.

The same position cannot be replicated by selling S & P stocks and buying FTSE 100 stocks.

The second paragraph emphasizes one convenience of the equity swap. Because it is based directly on an index, equity swap payoffs do not have any "tracking error." On the other hand, the attempt to replicate an index using underlying stocks is bound to contain some replication error.

3.1.2. Tax Advantages

Equity swaps are not only "cheaper" and more efficient ways of taking a position on indices, but may have some tax and ownership advantages as well. For example, if an investor wants to sell a stock that has appreciated significantly, then doing this through an outright sale will be subject to capital gains taxes. Instead, the investor can keep the stock, but, get into an equity swap where he or she pays the capital gains (losses) and dividends acquired from the stock, and receives some Libor-related return and a spread. This is equivalent to selling the stock and placing the received funds in an interbank deposit.

3.1.3. Regulations

Finally, equity swaps help in executing some strategies that otherwise may not be possible to implement due to regulatory considerations. In the following example, with the use of equity swaps investment in an emerging market becomes feasible.

Exmple

Since the Kospi 200 futures were introduced foreign securities houses and investors have been frustrated by the foreign investment limits placed on the instrument. They can only execute trades if they secure an allotment of foreign open interest first, and any allotments secured are lost when the contract expires. Positions cannot be rolled over. Foreigners can only hold 15% of the three-month daily average of open interest, while individual investors with "Korean Investor Ids" are limited to 3% Recognizing the bottle – neck, Korean securities houses such as Hyundai securities have responded by offering foreign participants equity swaps which are not limited by the restrictions.

The structure is quite simply. A master swap agreement is established between the foreign client and an off-shore subsidiary or a special purpose vehicle of the Korean securities company. Under the master agreement, foreigners execute equity swaps with the offshore entity which replicate the futures contract. Because the swap transactions involve two non-resident parties and are booked overseas, the foreign investment limits cannot be applied.

The Korean securities houses hedge the swaps in the futures market and book the trades in their proprietary book. Obviously, as a resident entity, the foreign investment limits are not applied to the hedging trade. Once the master swap agreement is established, the foreign client can contact the Korean securities company directly in Seoul, execute any number of trades and have them booked and compiled against the master swap agreement. (IFR, January 27, 1996)

The reading shows how restrictions on (1) ownership, (2) trading, and (3) rolling positions over, can be handled using an equity swap. The reading also displays the structure of the equity swap that is put in place and some technical details associated with it.

3.1.4. Creating Synthetic Positions

The following example is a good illustration of how equity swaps can be used to create synthetic positions.

Eexample

Equity derivative bankers have devised equity-swap trades to (handle) the regulations that prevent them from shorting shares in Taiwan, South Korea, and possibly Malaysia. The technique is not new, but has re-emerged as convertible bond (CB) issuance has picked up in the region, and especially in these three countries.

Bankers have been selling equity swaps to convertible bond arbitrageurs, who need to short the underlying shares but have been prohibited from doing so by local market regulations.

It is more common for a convertible bond trader to take a short equity – swap position with a natural holder of the stock. The stockholder will swap his long stock position for a long equity – swap position. This provides the CB trader with more flexibility to trade the physical shares. When the swap matures – usually one year later – the shares are returned to the institution and the swap is settled for cash. (IFR, December 5, 2001)

In this example a convertible bond (CB) trader needs to short a security by an amount that changes continuously. A convenient way to handle such operations, is for the CB trader to write an equity swap that pays the equity returns to an investor, and gets the investors' physical shares to do the hedging.

3.2. Using Interest Rate and Other Swaps

Interest rate swaps play a much more fundamental role than equity swaps. In fact, all swaps can be in balance sheet managements. Balance sheet contain several cash flows; using the swap, one can switch these cash flow characteristics. Swaps are used in hedging. They have zero value at time initiation and hence don't require any funding. A market practitioner can easily cover his positions in equity, commodities, and fixed income by quickly arranging proper swaps and then unwind these positions when there is no need for the hedge.

Finally, swaps are also *trading* instruments. In fact, one can construct *spread trades* most conveniently by using various types of swaps. Some possible spread trades are given by the following:

- Pay *n*-period fixed rate s_{t_0} and receive floating Libor with notional amount N.
- Pay L_{t_i} and receive r_{t_i} both floating rates, in the same currency. This is a basis swap.
- Pay and receive two floating rates in different currencies. This will be a currency swap.

As these examples show, swaps can pretty much turn every interesting instrument into some sort of "spread product." This will reduce the underlying credit risks, make the value of the swap zero at initiation, and, if properly designed, make the position relatively easy to value.

5.3 Two Uses of Interest Rate Swaps

We now consider two examples of the use of interest rate swaps.

5.3.1 Changing Portfolio Duration

Duration is the "average" maturity of a fixed-income portfolio. It turns out that in general the largest fixed-income liabilities are managed by governments, due to the existence of government debt. Depending on market conditions, governments may want to adjust the average maturity of their debt. Swaps can be very useful here. The following example illustrates this point.

Example

France and Germany are preparing to join Italy in using interest rate swaps to manage their debt. Swaps can be used to adjust debt duration and reduce interest rate costs, but government trading of over-the-counter derivatives could distort spreads and tempt banks to front-run sovereign positions. The U.S. and U.K. say they have no plans to use swaps to manage domestic debt.

As much as E150 bn of swap use by France is possible over the next couple of years, though the actual figure could be much less, according to an official at the French debt management agency. That is the amount that would be needed if we were to rely on only swaps to bring about "a [significant] shortening of the average duration of our debt," an official said. France has E644 bn of debt outstanding, with an average maturity of six years and 73 days.

The official said decisions would be made in September about how to handle actual swap transactions. "If E150bn was suddenly spread in the market, it could produce an awful mess," he said. (IFR, Issue 1392)

Using swap instrument, similar adjust to the duration of liabilities can routinely be made by corporation as well.

5.3.2 Technical Uses

Swaps have technical uses. The following example shows that they can be utilized in designing new bond futures contracts where the delivery is tied not to bonds, but to swaps.

Example

LIFFE is to launch its swap-based Libor Financed bond on October 18. Both contracts are designed to avoid the severe squeeze that has afflicted the Deutsche Terminboerse Bund future in recent weeks.

LIFFE's new contract differs from the traditional bind future in that it is swap-referenced rather than bond-referenced. Instead of being settled by delivery of cash bonds chosen from a delivery basket, the Libor Financed Bond is linked to the International Swap and Derivatives Association benchmark swap rate. At expiry, the contract is cash-settled with reference to this swap rate.

Being cash-settled, the Liffe contract avoids the possibility of a short squeeze- where the price of the cheapest to deliver bond is driven up as the settlement day approaches. And being referenced to a swap curve rather a bond basket, the contract eliminates any convexity and duration risk. The Libor Financial Bond replicates the convexity of a comparable swap position and therefore reduces the basis risk resulting from hedging with cash binds or bond futures.

An exchange-based contender for benchmark status, the DTB Bund, has drawn fire in recent weeks following a short squeeze in the September expiry. In the week before, the gross basis between the cheapest to deliver Bund and the Bund future was driven was driven down to-3.5.

The squeeze had been driven by a flight to quality on the back of the emerging market crisis. Open interest in the Bund future is above 600,000 contracts or DM15bn. In contrast, the total deliver basket for Germen government bond is roughly DM74 bn and the cheapest to deliver account to DM30 bn.

Official from the DTB have always contended that there will be no lack of deliverable Bunds. They claim actual delivery has been made in about 4% of open positions in the past. (IFR Issue 1327)

In fact, several new cash-settled futures contracts were recently introduced by LIFFE and CME on swaps. Swaps are used as the underlying instrument. Without the existence of liquid swap market futures contract would have no such reference point, and would have to be referenced to a bond basket.

4. Mechanics of Swapping New Issues

The swap engineering introduced in this chapter has ignored several minor technical points that need to be taken into account in practical applications. Most of these are minor, and, are due to differences in market conventions in bonds, money markets, and swap markets. In this section, we provide a discussion of some of these technical issues concerning interest rate swaps. In other swap markets, such as in commodity swaps, further technicalities may need to be taken into account. A more or less comprehensive list is as follows:

1. Real-world applications of swaps deal with new bond issues, and new bond issues imply fees, commissions, and other expenses that have to be taken into account in calculating the true cost of the funds. This leads to the notion of *all-in-cost*, which is different than the "interest rate" that will be paid by the issuer.
2. Interest rate swaps deal with fixed and floating rates simultaneously. The corresponding Libor is often taken as the floating rate, while the swap rate, or the relevant swap spread is taken as the fixed rate. Another real world complication appears at this point.
   Conventions for quoting money market rates such as Libor are quoted on a ACT/360 basis while some bonds are quoted on an annual or semi-annual 30/360 basis. In swap engineering, these cash flows are exchanged at regular times, and hence appropriate adjustments need to be made.
3. In this chapter we ignored credit risk. This greatly simplified the exposition because swap rates and corporate rates of similar maturities became equal. In financial markets, they usually are not. Issuers have different credit ratings and bonds sold by them carry credit spreads that are different from the swap spread. This gives rise to new complications in matching cash flows of coupon bonds and interest rate swaps. We need to look at some examples to this as well.
4. Finally, the mechanics of how new issues are swapped into fixed or floating rates and how this may lead to sub-Libor financing is an interesting topic by itself.

The discussion will be conducted with a real-life, *new issue*, explained next. First we report the "market reaction" to the bond, and second we have the details of the new issue.

Example

South korea's Shinhan Bank, rated Baal / BBB by Moody's and S & P, priced its US $ 200m three-year bond early last week (...). The deal came with a 4% coupon and offered a spread of 168.8bp over the two-year U.S. Treasury, equivalent to 63bp over Libor.

This was some 6bp wide of the Korea Development Bank (KDB) curve, although it was the borrower's intention to price flat to it. Despite failing to reach this target, the borrower still managed to secure a coupon that is the lowest on an Asian bond deal since the regional crisis, thanks to falling U.S. Treasury yields which have shrunk on a renewed flight to safe haven assets. (IFR, issue 1444)

Table 5-1 : Details of the New Issue

Shinhan Bank	
Amount	US$200m
Maturity	3 years (due July 2009)
Coupon	4%
Reoffer price	99.659
Spread at reoffer	168.8 bp over the two – year US Treasury
Launch date	July 23
Payment	July 29
Fees	20bp
Listing	London
Governing law	London
Negative pledge	Yes
Cross-default	Yes
Sales restrictions	US, UK, South Korea
Joint lead managers	ABN AMRO, BNP Paribas, UBS Warburg

Source: IFR issue 1444

Consider now the basic steps of swapping this new issue into floating USD funds. The issuer has to enter into a 3-year interest rate swap agreement. How should this be done, and what are the relevant parameters? Suppose at time of the issue the market makers were quoting the swap spreads shown in Table 5.2:

Table 5-2 : USD Swap index versus 12M Libor, Semi, 30/360F

Maturity	Bid Spread	Ask Spread	The Bid-ask Swap Rate
2 – Year	42	46	2.706 – 2.750
3 – Year	65	69	3.341 - 3.384
4 – Year	70	74	3.796 – 3.838
5 – Year	65	69	4.147 – 4.187
7 – Year	75	79	4.653 – 4.694
10 – Year	61	65	5.115 – 5.159
12 – Year	82	86	5.325 – 5.369
15 – Year	104	108	5.545 – 5.589
20 – Year	126	130	5.765 – 5.809
30 – Year	50	54	5.834 – 5.885

First we consider the calculation of all-in-cost for the preceding deal.

6.1. All-in-Cost

The information given in the details of the new issue implies that the coupon is 4%. But, this is not the true costs of funds from the point of view of the issuer. There are at least three additional factors that need to be taken into account. (1) The reoffer price is not 100, but 99.659. This means that for each bond, the issuer will receive less cash than the par. (2) Fees have to be paid. (3) Although not mentioned in the information in Table 5-1, the issuer has legal and documentation expenses. We assume that these were USD 75,000.

To calculate the fixed all-in-cost (30/360 basis), we have to calculate the *proceeds* first. Proceeds are the net cash received by the issuer after the sale of the bonds. In our case, using the terminology of table 5-1,

$$\text{Proceeds} = \text{Amount} \times \left(\frac{\text{Price}}{100} - \text{fees} \right) - \text{Expenses} \qquad (39)$$

Plugging in the relevant amount,

$$\text{Proceeds} = 200{,}000{,}000\,(0.99659 - 0.0020) - 75{,}000 \qquad (40)$$

$$= 198{,}843{,}000 \qquad (41)$$

Next, we see that the bond will make three coupon payments of 8 million each. Finally, the principal is returned in 3 years. The cash flows associated with this issue are summarized in Figure 5-11. What is the internal rate of return of this cash flow?

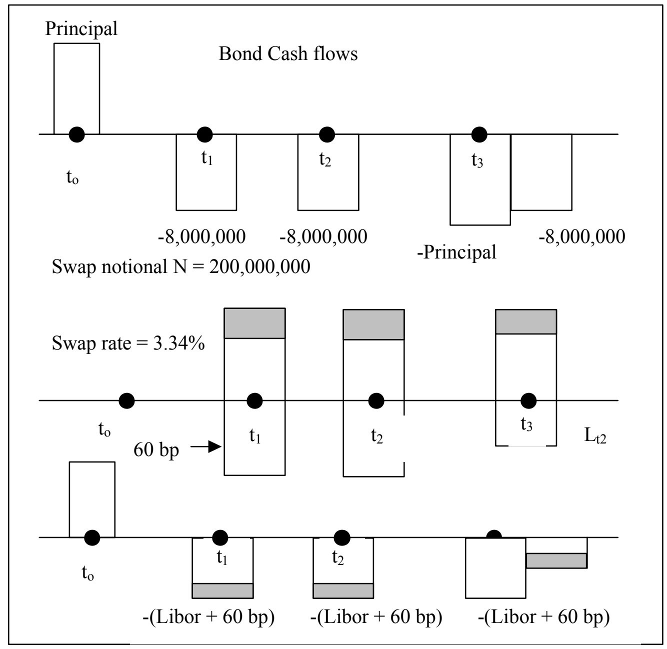


FIGURE 5-11

This is given by the formula

$$198{,}843{,}000 = \frac{8{,}000{,}000}{(1+y)} + \frac{8{,}000{,}000}{(1+y)^2} + \frac{8{,}000{,}000 + 200{,}000{,}000}{(1+y)^3} \qquad (42)$$

The y that solves this equation is the internal rate of return. It can be interpreted as the true cost of the deal, and it is the fixed all-in-cost under the (30/360) day – count basis. The calculation gives

$$y = 0.04209 \qquad (43)$$

This is the fixed all – in – cost.

The next step is to swap this issue into floating and obtain the floating all-in-cost. Suppose we have the same notional amount of $200 million and consider a fixed to floating 3-year interest rate swap. Table 5-2 gives the 3-year *receiver* swap rate as 3.341%. This is, by definition, a 30/360, semi-annual rate.

This requires converting the semi-annual swap rate into an annual 30/360 rate, denoted by r. This is done as follows:

$$(1+r) \quad = \quad (1+0.03341\ ½)^2 \qquad (44)$$

which gives

$$r = 3.369\% \qquad (45)$$

With a $200 million notional this is translated into three fixed receipts of

$$200{,}000{,}000\,(0.03369) = 6{,}738{,}000 \qquad (46)$$

each. The cash flows are shown in Figure 5-11b.

Clearly, the fixed swap receipts are not equal to the fixed annual coupon payments, which are $8 million each. Apparently, the issuer pays a higher rate than the swap rate due to higher credit risk. To make these two equal, we need to increase the fixed receipts by

$$8{,}000{,}000 - 6{,}738{,}000 = 1{,}262{,}000 \qquad (47)$$

This can be accomplished by increasing both the floating rate paid and the fixed rate received by equivalent amounts. This can be accomplished if the issuer accepts paying Libor *plus* a spread equivalent to the 66bp. Yet, here the 66bp is p.a. 30/360, whereas the Libor convention is p.a. act/360. So the basis point difference of 66 bp may need to be adjusted further. The final figure will be the floating all-in-cost and will be around 60bp.

6.2. Another Example

Suppose there is an A-rated British entity that would like to borrow 100m sterling (GBP) for a period of 3 years. The entity has no preference toward either floating or fixed-rate funding, and intends to issue in Euromarkets. Market indicates that if the entity went ahead with its plans, it could obtain fixed-rate funds at 6.5% annually. But the bank recommends the following approach.

It appears that there are nice *opportunities* in USD – GBP currency swaps, and it makes more sense to issue a floating rate Eurobond in the USD sector with fixed coupons. The swap market quotes funding at Libor + 95 bp in GBP against USD rates for this entity. Then the proceeds can be swapped into sterling for a lower all –in-cost. How would this operation work? And what are the risks?

We begin with the data concerning the new issue. The parameters of the newly issued bond are as follows.

Table 5-3 : The New Issue

Amount	USD100 million
Maturity	2 years
Coupon	6% p.a.
Issue price	100 ¾
Options	none
Listing	Luxembourg
Commissions	1 ¼
Expenses	USD 75000
Governing law	English
Negative pledge	Yes
Pari passu	Yes
Cross default	Yes

Now, the issuer would like to swap these proceeds to floating rate GBP funds. In doing this, the issuer faces the following market conditions:

Table 5-4 : Swap Market Quotes

Spot exchange rate GBP-USD	1.6701/1.6708
GBP 2-year interest rate swap	5.46/51
USD-GBP currency swap	+4/-1

We first work out the original and the swapped cash flows and then calculate the all-in-cost, which is the real cost of funds to the issuer after the proceeds are swapped into GBP.

The first step is to obtain the amount of cash the issuer will receive at time t_o and then determine how much will be paid out at t_1, t_2. To do this, we again need to calculate the *proceeds* from the issue.

- The issue price is 100.75 and the commissions are 1.25%. This means that the amount received by the issuer before expenses, is

$$\frac{(100.75 - 1.25)}{100} \quad 100{,}000{,}000 = 99{,}500{,}000 \tag{48}$$

We see that the issue is sold at a premium which increases the proceeds, but once commissions are deducted, the amount received falls below 100 million. Thus, expenses must be deducted

$$\text{Proceeds} = 99{,}500{,}000 - 75{,}000 = 99{,}425{,}000 \tag{49}$$

Given the proceeds, we can calculate the effective cost of fixed rate USD funds for this issuer. The issuer makes two coupon payments of 6% (out of the 100 million) and then pays back 100 million at maturity. At t_0, the issuer receives only 99,425,000. This cash flow is shown in Figure 5-12. Note that unlike the theoretical examples, the principal paid is not the same as the principal received. This is mainly due to commissions and expenses.

- From this cash flow we can calculate the internal rate of return y_{to} by solving the equation

$$99{,}425{,}000 = \frac{60{,}000}{(1+y)} + \frac{60{,}000}{(1+y)^2} + \frac{100{,}000{,}000}{(1+y)^3} \tag{50}$$

The solution is

$$y_{to} = 6.3150\% \tag{51}$$

Hence, the true fixed cost of USD funds is greater than 6%.

The issuer will first convert this into floating rate USD funds. For this purpose, the issuer will sell a swap. That is to say, the issuer will receive fixed 5.46% and pay floating Libor flat. This is equivalent to paying approximately USD Libor + 54bp. Finally, the issuer will convert these USD floating rate funds into GBP floating rate funds by paying floating GBP and receiving floating USD.

3. Some Conventions

If you have a coupon bond and the payment date falls on a nonworking day, then the payment will be made on the first following working day. But the amount does not change. In Swaps, this convention is slightly different. The payment is again delayed to the next working day. But, the payment amount will be adjusted according to the actual number of days. This means that the payment dates and the amounts may not coincide exactly in case swaps are used as hedges for fixed-income portfolios.

3.1. Quotes

Suppose we see quotes on interest rate swaps or some other liquid swap. Does this mean we can deal on them? Not necessarily. Observed swap rates may be available as such only to a bank's best customers; others may have to pay more. In practice, the bid-ask spreads on liquid instruments are very tight, and a few large institutions dominate the market.

8. Currency Swaps Versus FX Swaps

We will now compare currency swaps with FX-swaps introduced in Chapter 3. A Currency swap has the following characteristics:

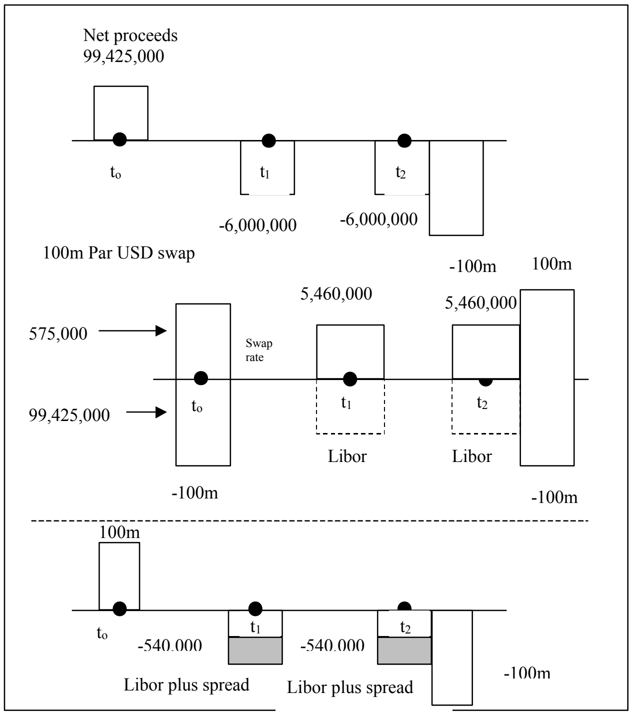


Figure 5-12

1. Two principals in different currencies and of equal values are exchanged at the start date t_o.
2. At settlement dates, interest will be paid and received in different currencies, and according to the agreed interest rates.
3. At the end date, the principals are re-exchanged at the same exchanged rate.

A simple example is the following. 100,000,000 euros are received and against these 100,000,000 e_{to} dollars are paid, where the e_{to} is the "current" EUR/USD exchange rate. Then, 6-month Libor-based interest payments are exchanged twice. Finally, the principal amounts are exchanged at the end date at the same exchanged rate e_{to}, even though the actual exchanged

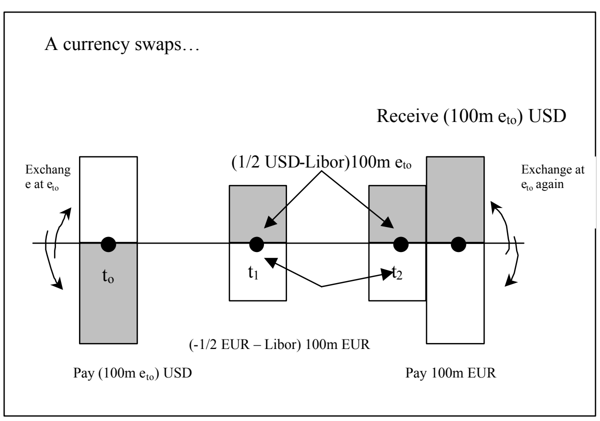


Figure 5-13

rate e_{t2} at time t_2 may indeed be different than e_{to}. See Figure 5-13.

The FX-swap for the same period is in Figure 5-14. Here, we have no interim interest payments, but instead the principals are re-exchanged at a different exchange rate equal to

$$F_{to} = e_{to} \frac{1+ L_{to}^{USA}d}{1+ L_{to}^{USA}d} \tag{52}$$

Why this difference? Why would the same exchange rate be used to exchange the principals at start and end dates of a currency swap while different exchange rates are used for an FX-swap?

We can look at this question from the following angle. The two parties are exchanging currencies for the period of 1 year. At the end of the year they are getting back their original currency. But during the year, the interest rates in the two currencies would normally be different. This difference is explicitly paid out in the case of currency swaps during the life of the swaps as interim interest payments. As a result, the counterparties are ready to receive the exact original amounts back. The interim interest payments would compensate them for any interest rate differentials.

In the case of FX-swaps, there are no interim interest payments. Hence, the compensation must take place at the end date. Thus, the interest payments are bundled together with the exchange of principals at the end date.

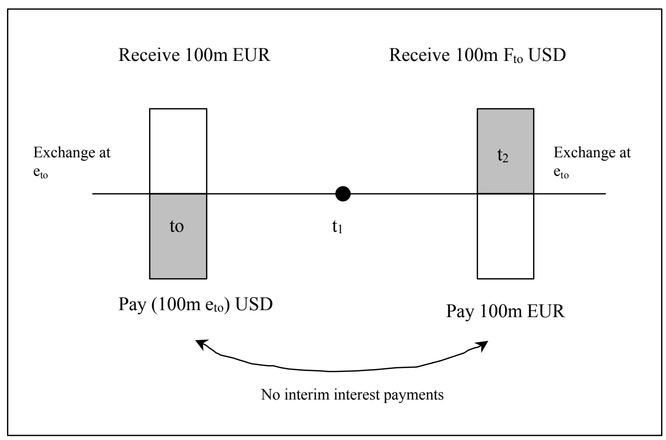


Figure 5-14

8.1 Another Difference

Looked at from a financial engineering perspective, the currency swap is like an exchange of two FRNs with different currencies and no credit risk. The FX-swap, on the other hand, is like an exchange of two zero-coupon bonds in different currencies.

Because the Libor rates at time t_1 are unknown as of time t_o, the currency swap is subject to slightly different risks than FX-swaps of the same maturity.

REPO MARKET STRATEGY IN FINANCIAL ENGINEERING

Objectives

- After studying this lesson you will be able to understand repo market as a complementary to and or an alternative to swap.

Friends,

this is a nontechnical lesson which deals with a potentially confusing operation. The chapter briefly reviews repo markets and some uses of repo. This is essential for understanding many standard operations in financial markets.

Many financial engineering strategies require the use of the repo market. The repo market is both a complement and an alternative to swap markets. During a swap transaction, the market practitioner conducts a simultaneous "sale" and "purchase" of two sequences of cash flows generated by two different securities. For example, returns of an equity instrument are swapped for floating rate Libor. This is equivalent to selling the equity, receiving cash, and then buying a floating rate note (FRN). These operations are combined in an equity swap and accomplished without actually buying and selling the involved assets or exchanging the original principals. With no exchange of cash, flexible maturities, and liquid markets, swaps become a fundamental tool. Using swaps, a complex sequence of operations can be accomplished efficiently, quickly, and with little risk.

Repo transactions provide similar efficiencies, with two major differences. In swaps, the use of cash is minimized and the ownership of the instruments does not change. In a repo transaction both cash and ownership changes hands. Suppose a practitioner does need cash or needs to own a security. Yet, he or she does not want to give up or assume the ownership of the security permanently. Swaps are of no help, but a repo is.

Repo is a tool that can provide us cash without requiring the sale, or giving up eventual ownership of the involved assets. Alternatively, we may need a security, but we may not want to own this security' permanently. Then we must use a tool that secures ownership, without really requiring the purchase of the security. In each case, these operations require either a temporary use of cash or a temporary ownership of securities. Repo markets provide tools for such operations. With repo transactions, we can "buy" without really buying, and we can "sell" 'without really selling. This is similar to swaps in a sense, but most repo transactions involve exchanges of cash or securities, and this is the main difference with swap instruments.

In each case, the purpose behind these operations is not "long-term." Rather, the objective is to conduct daily operations rather smoothly, take directional positions, or hedge a position more efficiently.

2. What Is Repo?

We begin with the standard definition. A repo is a repurchase agreement where a repo dealer sells a security to counterparty and simultaneously agrees to buy it back at a predetermined price and; it a predetermined date. Thus, it is a sale and a repurchase written on the same ticket. In a repo, the dealer first delivers the security and receives cash from the client. If the operation is reversed- that is to say, the dealer first buys the security and simultaneously sells it back at a predetermined date and time - the operation is called a reverse repo, or is simply referred to as reverse.

At first glance, the repo operation looks like a fairly simple transaction that would not contribute to the methodology of financial engineering. This is not true. In fact, in terms of practical applications of financial engineering repo may be as common as swaps.

Consider the following experiment. Suppose an investor wants to buy a security using short -term funding. If he borrows these funds from a bank and then goes to another dealer to buy the bond, the original loan will be nonsecured. This implies higher interest costs. Now, if the investor uses repo by buying first, and then repoing the security, he can get the needed funds cheaper because there will be a collateral behind the "loan," As a result, both the transaction costs and the interest rate will be lower. In addition, in the repo operation, transactions are grouped and written on single ticket. Given the lower risks, higher flexibility, and other conveniences, repo transactions are very liquid and practical.

With a repo the sequence of transactions changes. In a typical outright purchase a market professional would

Secure funds $\rightarrow$ Pay for the security $\rightarrow$ Receive the security (1)

When repo markets are used for buying a security, the sequence of transactions becomes:

Buy the bond $\rightarrow$ immediately repo it out $\rightarrow$ Secure the funds $\rightarrow$ Pay for the bond (2)

In this case, the repo market is used for finding cheap funding for the purchases the practitioner needs to make. The bond is used as collateral. If this is a default-free security, borrowed funds will come with a relatively low repo rate.

Similarly, shorting securities also becomes relatively easy. The market participant will use the repo market and go through the following steps:

Deliver the cash and borrow the bond $\rightarrow$ Return the bond and receive cash plus interest (3)

The market practitioner will earn the repo rate while borrowing the bond. This is equivalent to the market practitioner holding a short-term bond position. The bond is not purchased, but it is "leased".

2.1. A Convention

The following can get very confusing if not enough attention is paid to it. In repo markets terminology is set from the point of view of the repo dealer. Also, words such as "borrowing" and "lending" are used as if the item that changes hands is not cash, but a security such as a bond or equity. In particular, the terms "lender" or "borrower" are determined by the lending and borrowing of a security and not of "cash" - although in the actual exchange, cash is changing hands.

Accordingly, in a repo transaction where the security is first delivered and cash is received, the repo dealer is the "lender" - he or she lends the security and gets cash. This way, the repo dealer has raised cash. If, on the other hand, the same operation was initiated by a client, and the counterparty was a repo dealer, the deal becomes a reverse repo. The dealer is borrowing the security, the reverse of what happens in a repo operation.

2.2. Special Versus General Collateral

Repo transactions can be classified into two categories. Sometimes, specific securities receive special attention from markets. For, example, some bonds become cheapest-to-deliver. The "shorts" who promised delivery in the bond futures markets are interested in a particular bond and not in others that are similar. This particular bond becomes very much in demand and goes special in the repo markets. A repo transaction that specifies the particular security in detail is called a special repo. The security remains special as long as the relative scareity persists in the market.

Otherwise, in a repo deal, the party that lends the securities can lend any security of a similar risk class. This type of security is called general collateral. One party lends U.S. government bonds against cash, and the counterparty does not care about the particular bonds this basket contains. Then the collateral could be any Treasury bond.

The special security will have a higher price than its peers, as long as it remains special. This means that to borrow this security, the client gives up his or her cash at a lower interest rate. After all, the client really needs this particular bond and will therefore have to pay a "price." The price is agreeing to a lower repo rate.

The interest rate for general collateral is called the repo rate. Specials command a repo rate that is significantly lower. In this latter case, the cash can be re-lent at a higher rate via a general repo. The original owner of the "special" benefits.

Example

Suppose repo rate quotes are 4.5% to 4.6%. You own a bond worth 100, which by" chance goes special the next day. You can lend your bond for say USD 100 and get cash for 1 week and pay only 2.5%. This is good, since you can immediately repo this sum against general collateral and earn an annual rate of 4.5% on the 100. You have earned an enhanced return on your bond because you just happened to hold something special.

When using bond market data in researeh, it is important to take into account the existence of specials in repo transactions. If "repo specials" are mixed with transactions dealing with general collateral, the data may exhibit strange variations and may be quite misleading. This point is quite relevant since about 20% of repo transactions involve specials.

2.2.1. Why Do Bonds Go Special?

There are at least two reasons why some securities go special systematically. For one, some bonds are cheapest-to-deliver (CTD) in bond futures trading (see the case study at the end of this chapter). The second reason is that on-the-run issues are more liquid and are therefore more in demand by traders in order to support hedging and position-taking activities. Such "benchmark" bonds often go special. This is somewhat paradoxical, as the more liquid bonds become more expensive to obtain relative to others.

As an example, consider the so-called butterfly trades in the fixed-income sector. Nonparallel shifts that involve the belly of the yield curve are sometimes called butterfly shifts. These shifts may have severe implications for balance sheets and fixed-income portfolios. Trader's use2-5-10 year on-the-run bonds to put together hedging trades, to guard, or speculate against such yield "curve movements. These trades are called butterfly trades. The on-the-run bonds used in such strategies may become "benchmarks" and may go special.

2.3. Summary

We can now summarize the discussion. What are the advantages of repo transactions?

1. A repo provides double security when lending cash. These are the (high) credit rating of a repo dealer and the collateral.
2. A "special" repo is a unique and convenient way to enhance returns.
3. By using repo markets, traders can short the "market and raise funding efficiently. This improves general market efficiency and trading.
4. Financial strategies and product structuring will benefit due to lower transaction costs more efficient use of time, and lower funding costs.

We now consider various types of repo or repo-type transactions.

3. Types of Repo

The term "repo" is used for selling and then simultaneously repurchasing the same instrument. But in practice, this operation can be done in different ways, and these lead to slightly different repo categories.

3.1. Classic Repo

A classic repo, is also called a U.S. - style repo. This is the operation that we just discussed. A repo dealer owns a security that he or she sells at a price, 100. This security he or she immediately promises to repurchase, at 100, say in I month. At that time, the repo dealer returns the original cash received, plus the repo interest due on the sum.

Example

An investor with a fixed income portfolio wants to raise cash for a period of one week only. This will be done through lending a bond on the portfolio. Suppose the trade date is Monday morning. The parameters of the deal are as follows:

Value date: Deal date + 2 days

Start proceeds: 50 million euro

Collateral: 6-3/4% 4/2003 Bund (the NOMINAL value equals 47.407m)

[1]An on-the-run issue is the latest issue for a particular maturity, in a particular risk class. For example, an on-the-run 100year treasury will be the last l0-year bond sold in a treasury action. Other to-year bonds will be off-the-run.

Term: 7 days

Repo rate: 4.05%

End proceeds: Start proceeds + (start proceeds X repo rate X term)

This gives

EUR 50m + (EUR 50m X .0405 X 7/360) = EUR 50 039375

Repo interest: 39375

Thus, by lending 47 407 000 of nominal bonds (DBRs), the investor borrows EUR 50 million. This situation is shown in Figure 6-1.

The difference between the nominal and 50m is due to the existence of accrued interest. Accrued interest needs to be added to the nominal. That is to say the calculations are done using bond's dirty price.

Before we look at further real-life examples, we need to consider other repo types.

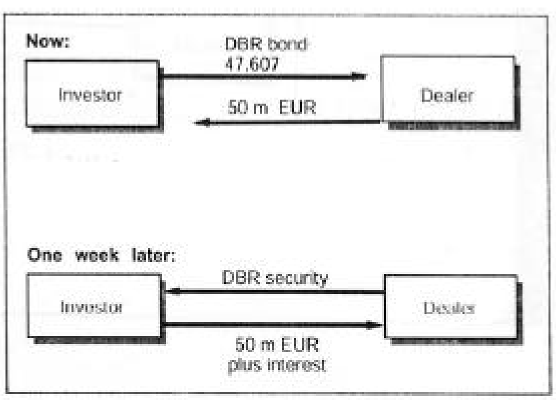
3.2. Sell and Buy, Back,

A second type of repo is called sell and buy-back. The end result of a sell and buy-back is not different from the classic repo. But, the legal foundations differ, which means that credit risks may also be different. In fact, sell and buy-backs exist in two different ways. Some are undocumented. Two parties write two separate contracts at the same time to. One contract involves a spot sale of a security, while the other involves a forward repurchase of the same security at a future date. Everything else being the same, the two prices should incorporate the same interest component as in the classic repo. In the documented sell and buy-back, there is a single contract, but the two operations are again treated as separate.

Example

We use the same parameters as in the previous example, but the way we look at the operation is different although the interest earned is the same:

Nominal: EUR 47.607 million Bund 63/4% 4/2003



Start price: 101.971

Plus accrued interest: 3.05625

Total price: 105

Start proceeds: EUR 50,000,322.91

End price: 101.922459

Plus accrued interest: 3.1875

Total price 105.109959

End proceeds 50 039 698.16

Repo interest earned: EUR 50000322.91 X .0405 X 7/360 = 39375

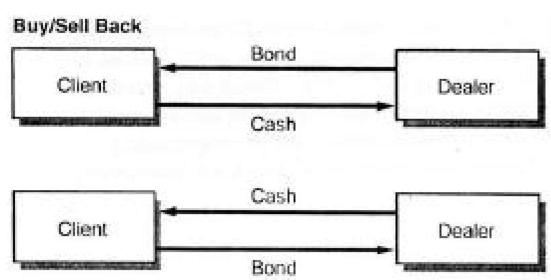
In this case the investor's interest cost in the difference between the purchase price and selling price. The interest earned is exactly the same as in classic repo, but the way interest rate is characterized is different. We show the deal in Figure 6-2.

The major difference between the two repo types lies not in the mechanics, or in interest earned, but in the legal and risk management aspects. First of all, sell and buy-backs have no mark-to-market. So they are "easier" to book. Second, in case of undocumented sell and buy-backs, no documentation means lower legal expenses and lower administrative costs. Yet, associated credit risks may be higher. In particular, with s0ell and buy-backs there is no specific right to offset during default.

3.3. Securities Lending

Securities lending is older than repo as a transaction. It is also somewhat less practical than repo. However, the mechanics of the operation are similar. The main difference is that, one of the parties to the transaction may not really need the cash that a repo would generate. But, this party may still want to earn a return, hence, the party simply lends out the security for a fee. Any cash received may be deposited as collateral with another entity.

Clearing firms such as Euroclear and Cedel do securities lending. Suppose a bond dealer is a member of Cedel. The dealer sold a bond that he or she did not own, and could not find in the markets for an on-time delivery. This may result in failure to deliver. Cedel can automatically lend this dealer a security, by borrowing (at random) from another member.



Notice that, here securities can be lent not only against cash; but against other securities as well. The reason is simple: the lender of the security does not need cash, but rather needs collateral. The collateral can even be a letter of credit or any other acceptable form.

One difference between securities lending and repo is in their quotation. In securities lending, a fee is quoted instead of a repo rate.

Example

Nominal: GBP 10 million 8.5% 12/07/05 is lent for 2 weeks

Collateral: GBP /0.62 million 8% 10/07/06

Fee: 50bp

Total Fee: 50bp X (/4days/360) x GBP 10 million

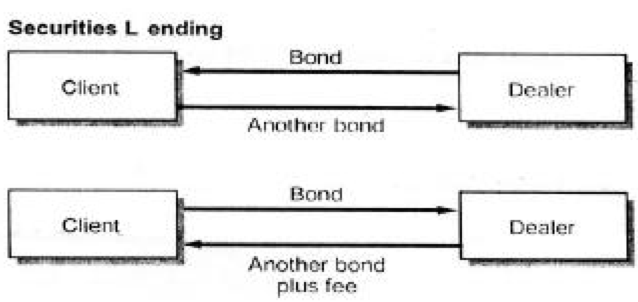
Obviously, the market value of the collateral will be at least equal to the value of borrowed securities. All other terms of the deal will be negotiated depending on the credit of the borrowing counterparty and the term. This transaction is shown in Figure 6-3.

3.4. Custody and Repo Types

There are different ways of holding the collateral. A classic type is delivery-repo. Here, the security is delivered to the counterparty. It is done either by physical delivery or as a transfer of a book entry. A second category is called hold-in-custody repo, where the "seller" (lender) keeps the security on, behalf of the buyer, during the term of the repo. This is, either because it is impossible to make the transfer or because it is not worth it, due to time or other limitations.

The third type of custody handling is through a triparty repo, where a third party holds the collateral on behalf of the "buyer" (borrower). Often the two parties have accounts with the, same custodian. Then the triparty repo involves simply a transfer of securities from one account to another. This will be cheaper, since, fewer fees or commissions are paid. In this case, the custodian also handles the technical details of the repo transaction such as (1) ensuring that delivery versus payment is made and (2) ensuring marking to market of the collateral.

According to all this, a good clearing, custody, and settlement infrastructure is an essential prerequisite for a well-functioning repo market.



3.4.1. What Is a Matched-Book Repo Dealer?

Repo dealers are in the business of writing repo contracts. At any time, they post bid and ask repo rates for general, as well as, special collateral. In a typical repo contributor page of Reuters or Bloomberg, the specials will be clearly indicated as special and will command special prices (i.e., special repo rates). At any time, the repo dealer is ready to borrow and lend securities, whether they are special or general collateral. This way, books are "matched." But this does not mean that dealers don't take one-way positions in the repo book. Dealers' profit comes, from bid-ask spreads and from taking market exposure when they think it is appropriate to do so.

3.5. Aspects of the Repo Deal

We briefly summarize some further aspects of repo transactions.

1. A repo is a temporary exchange of securities against cash. But, it is important to realize that the party who borrows the security has temporary ownership to the security. The underlying security can be sold. Thus, repo can be used for short-selling.
2. Because securities borrowed through repo can, in general, be sold, the securities returned in the second leg of the repo do not have to be identical. They can be "equivalent," unless specified otherwise in the repo deal.
3. In a repo deal, the lender of the security is transferring the title for a short period of time. But, the original owner is keeping the risk and the return associated with the security. Thus, coupon payments due during the term of the repo are passed on to the original owner of the security.

   The second reason that the risk remains with the original owner is related to the marking to market of the borrowed securities. For example, during the term of the repo, markets Q1ight crash, and the value of the collateral may decrease. The borrower of the security then has the right to demand additional collateral. If the value of the securities. Increases, some of the collateral has to be returned.
4. Coupon or dividend payments during the term of the repo are passed on to the original owner. This is called manufactured dividend, and can occur at the end of the repo deal or some time during the term of the repo.[2]
5. Repo markets have a practice similar to that of initial margin in futures markets. It is called, haircut. The party borrowing the bonds may demand additional security for delivering cash. For example, if the current market value of the securities is 100, the party may pay only 98 against this collateral. Note that, if a client faces a 2% haircut when he or she borrows cash in the repo market, the repo dealer can repo the same security with zero haircut and benefit from this transaction.
6. In the United States and the United Kingdom, repo documentation is standardized. A standard repo contract is known as a PSA/ISMA global repo agreement.
7. In the standard repo contract, it is possible to substitute other securities for the original collateral, if the lender desires so.
8. As mentioned earlier, the legal title of the repo passes on to the borrower (in a classic repo), so that in case of default, the security automatically belongs to the borrower (buyer). There will be no need to establish ownership.

2 Manufactured dividend is due on the same date as the date of the coupon. But for sell and buy-back this changes, the coupon is paid at the second leg.

Finally we should mention that settlement in a classic repo will be delivery versus payment (DVP). For international securities, the parties will in general use Euroclear and Cedel.

There are three possible ways to settle repo transactions. First there is cash settlement, which involves the same-day receipt of "cash." Second, there is regular settlement. The cash will be, received on the first business date following a trade date. Third we can have a skip settlement. Here, cash will be received 2 business days after the trade date.

4. Equity Repos

If we can repo bonds out, can we do the same with equities? This would indeed be very useful. Equity repos are becoming more popular, but, from a financial engineering perspective, there, ' are potential technical difficulties:

1. Equities pay dividends and make rights issues. There are mergers and acquisitions. How would we take these events into consideration in a repo deal? It is easy to account for coupons because these are homogeneous payouts. But, mergers, acquisitions, and rights issues imply much more complicated changes in the underlying equity.
2. It is relatively easy to find 100 million USD of a single bond to repo out; how do we proceed with equities? To repo equities worth 100 million USD, a portfolio needs to be put together.This complicates the instrument, and makes it harder to design a liquid contract.
3. The non-existence of a standard equity repo agreement also hampers liquidity. In the UK, this business is conducted with an equity annex to the standard repo agreement.
4. Finally, we should remember that equity has higher volatility. This implies more frequent, marking to market.

We should also point out that, some investment houses carry old-fashioned equity swaps and equity loans, and then label them as equity repos.

5. Repo Market Strategies

The previous sections dealt with repo mechanics and terminology. In this section, we start using repo instruments to devise financial engineering strategies.

5.1. Funding a Bond Position

The most classic use of repo is in funding fixed-income portfolios, A dealer thinks that it is the right time to buy a bond. But, as is the case for market professionals, the dealer does not have cash in hand. Then, he can use the repo market. A bond is bought and repoed out at the same time to secure the funds needed to pay for it. The dealer earns the bond return. His cost will be the repo rate.

The same procedure may be used to fund a fixed-income portfolio and to benefit from any opportunities in the market, as the following reading shows.

Example

Foreign fund mO1tagers have recently been putting on bond versus swap spread plays in the Singapore dollar-denominated market to take advantage of an expected widening in the spread between the term repo rate and swap spreads. "It's one of the oldest trades in the book," said [a trader] noting that it has only recently become feasible in the local market.

In a typical trade, an investor buys 10-year fixed-rate Singapore government bonds yielding 3.58%. and then raises cash on these-bonds via the repo market and pays an annualized funding rate of 2.05%. . . . At the same time the investor enters a la-year interest-rate swap in which it pays 3.715% fixed and receives the floating swap offer rate, currently 2.31 %. While the investor is paying out 13.5 basis points on the difference between the bond yield it. receives and the fixed rate it pays in the swap, the position makes 26 bps on the spread between the floating rate the investor receives in the swap and the term repo funding rate, he explained. The absolute levels of the repo and swap; offer rate may change, but the spread between them is most likely to widen, increasing the profitability of the transaction.

One of the most significant factors that has driven liquidity in the repo is that in the last few months the Monetary Authority of Singapore has started using the repo market for monetary authority intervention, rather than the foreign exchange market which it had traditionally used. (Based on an article in Derivatives Week).

We will analyze this episode in detail, using the financial engineering tools developed in earlier chapters. For simplicity, we assume that the underlying bonds and the swap have 3-year maturities with the numerical values given in the example above.3 The bond position of the trader is shown in Figure 6-4a. A price of 100 is paid at t_o to receive the coupons and the principal. Figure 6-4b shows the swap position. The swap "hedges" the fixed coupon payments, and "converts" the fixed coupon receipts from the bond into floating interest receipts. The equivalent of Libor in Singapore is Sibor. After the swap, the trader is receiving Sibor-13.5 bp. This is shown in Figure 6-4c which adds the first two cash flows vertically. At this point, we see another characteristic of the position: The trader IS receiving the floating payments, but, still has to make the initial payment of 100. This means the trader has to get these funds from somewhere.

One possibility is to borrow them from the market. A better way to obtain them is the repo. By lending the bond as collateral, the player can get the needed funds, 100 - assuming zero haircut. This situation is now shown in Figure 6-5. We consider, artificially, a 1-year repo contract and .a_sume tbat the repo can be rolled over at unknown repo rates R_{tl} and R_{t2} in future periods. According to the reading, the current repo rate is known:

$$R_{t0} = 2.05\% \tag{4}$$

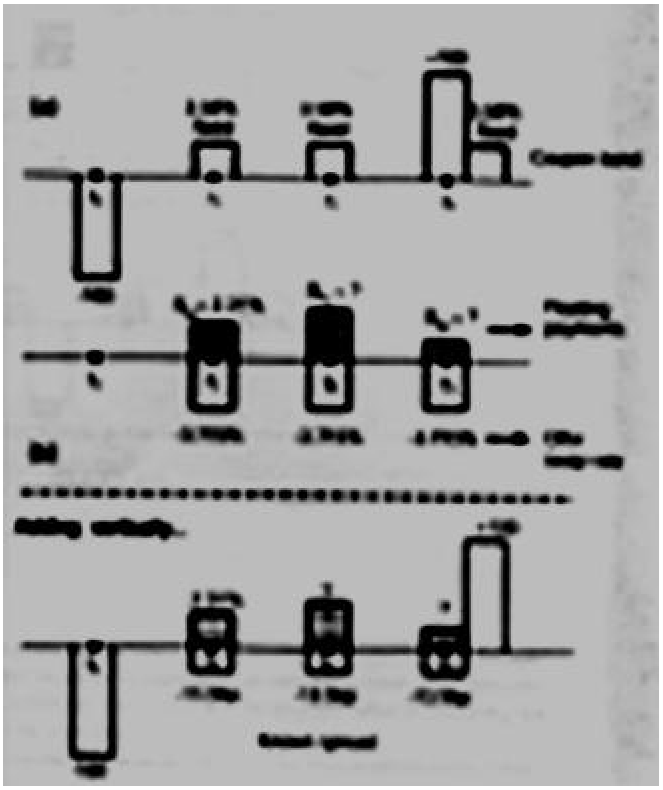
Adding the first two positions in Figure 6-5 vertically, we obtain the final exposure of the market participant.

The market participant has a 12.5 bp net gain for I year. But, more importantly, the final position has the following characteristic. The figure shows that the market participant is long a forward floating rate bond, which pays the floating Sibor rates S_{tl} and S_{t2} with the following expectation:

$$S_{tl} > R_{tl} + 13.5 \text{ bp} \tag{5}$$

$$S_{t2} > Rt2 + 13.5 \text{ bp} \tag{6}$$

That is to say, if the spread between future repo rates and Sibor tightens below 13.5 bp, the position will be losing money. This is the risk implied by the overall position. The lower part of



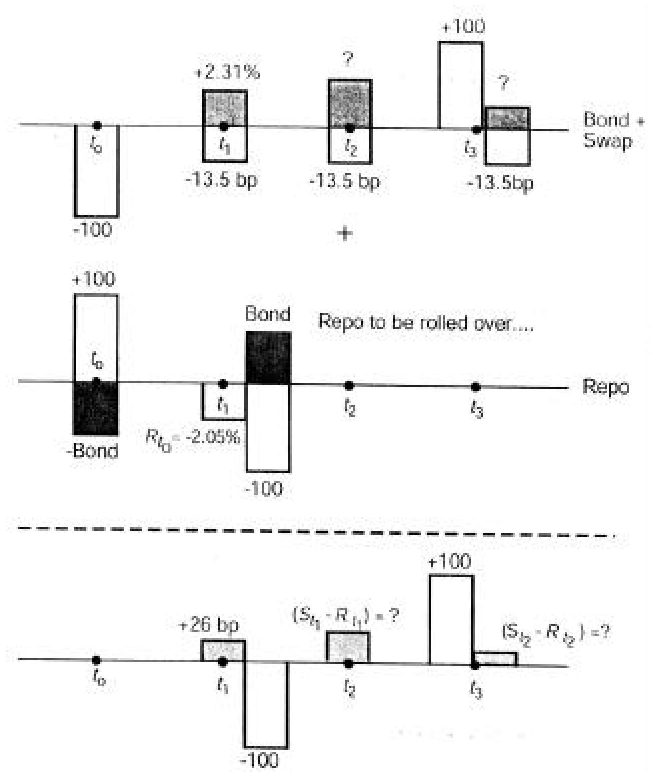


Figure 6-5

figure 6-5 shows how this exposure can be hedged. To hedge the position, we would need to go short the same bond forward.

5.1.1. Risks and Pricing Aspects

The position studied in the previous section is quite common in financial markets. Practitioners call these arbitrage plays or just arb. But, it is clear from the cash flow diagrams that this is not the arbitrage that an academic would refer to. In the preceding example there was no initial investment. The immediate net gain was positive, but the practitioner had an open position and this was risky. The position was paying net 12.5 bp today, however, the trader was taking the risk that the future spreads between repo rates and Sibor could tighten below 13.5 basis points. It is true that a 6-month Sibor has a longer tenor than, say, a I-month repo rate and hence assuming, positively sloped yield curve, the spread will be positive, but this cannot be guaranteed.

Second, the player is assuming different credit risks. He or she is paying a low 2.05 % on the repo financing because it is backed by Singapore government bonds. On the other hand, the 2.31 % received from the Sibor side is on a loan made to a high-quality private sector credit. Thus, the question remains. Is the net return of 12.5 bp worth the risks taken?

5.1.2. An Arbitrage Approach

There is a way to evaluate the appropriateness of the 12.5 bp return mentioned in the example. In fact, the market practitioner's final position is equivalent to owning a basis swap between the repo rate and the floating swap reference rate. After all, the position taker is receiving the floating rate in the swap and paying the repo rate.

Suppose the repo and swaps have identical settlement dates t_i. The final position is one where, at each settlement date, the position taker will receive

$$(Lt_{i-1} + 12.5\text{ bp} - R_{t_{i-1}})\, d\, N \qquad (7)$$

Clearly, this is similar to the settlement of a basis swap with a 12.5 bp spread and notional amount N. If such basis swaps traded actively in the Singapore market, one could evaluate the strategy by comparing the net return of 12.5 bp with the basis swap spread observed in the market. In case they are equal, then the same position can be taken directly in the basis swap market. Otherwise, if the basis swap rate is different than 12.5 bp, then a true arbitrage position may be put in place by buying the cheaper one and simultaneously selling the more expensive position.

5.2. Futures Arbitrage

Repo plays a special role in bond and T-bill futures markets. Consider a futures position with expiration to + 30 days. In 30 days, we will take possession of a default-free zero coupon bond with maturity T at the predetermined futures price P_{t_0}.

Hence, at settlement, P_{to} dollars will be paid and the I-year bond will be received. Of course, at to + 30 the market value of the bond, will be given by $B(t_o + 30, T)$ and will, in general, be different than the contracted P_{to}.

The repo market can be used to hedge this position. This will lead us to the important notion of implied repo rate.

How can we use repo to hedge a short bond futures position? The idea has been dealt with earlier, while discussing cash and carry trades. Secure funding, and buy a T + 30 day maturity bond at to. When time $t_o + 30$ arrives, the maturity left on this bond will be T, and thus the cash and carry will result in the same position as the futures. The practitioner borrows USD at t_o buys the $B(t_0, T + 30)$ bond, and keeps this bond until time $t_0 + 30$.

The novelty here is that we can collapse the two steps into one, by buying the bond and then immediately repoing it to secure financing. The result should be a futures position with an equivalent price.

This means that the following equation must be satisfied:

$$P_{to} = B(t_o, T + 30)\left(1 + R_{to}\frac{30}{360}\right) \quad (8)$$

In other words, once the carry cost of buying the T + 3O-day maturity bond is included the, total amount paid should equal P_{t0} the futures price of the future price of the bond.

Given the market quotes on the P_{to}, $B(t_o, T+30)$, market practitioners solve for the unknown

R_{to} and call this the implied repo rate:

$$R_{to} = \left(\frac{P_{to}}{B(t_0, T + 30)}\right) 30 \; \frac{360}{} \quad (9)$$

The implied repo rate is a pure arbitrage concept and shows the carry cost for fixed-income dealers.

5.3. Hedging a Swap

Repo can also be used to hedge swap positions. Suppose a dealer transacts a 100 million 2-year swap with a client. The dealer will pay the fixed 2-year treasury plus 30 bp, which brings the bid swap rate to, say, 5.95%. As usual, Libor will be received. The dealer hedges the position by buying a 2-year treasury.

In doing this, the dealer expects to transact another 2-year swap "soon" with another client, and receives the fixed rate. Given that the asking rate is higher than the bid swap rate, the dealer will capture the bid-ask spread. Suppose the ask side swap spread is 33 bp. ‘

Where does the repo market come in? The dealer has hedged the swap with a 2- year treasury, but how is this treasury funded? The answer is the repo market. The dealer buys the treasury, and then immediately repos it out overnight. The repo rate is 5.61 %. The dealer expects to find, a matching order in a few days. During this time, the trader has exposure to, (I) changes in the. : swap spread and (2) changes in the repo rate.

5.4. Tax Strategies

Consider the following situation:

- Domestic bondholders pay a withholding tax, while foreign owners don't. Foreign investors receive the gross coupons.

The following operation can be used. The domestic bondholder repos out the bond just before the coupon payment date to a foreign dealer (i.e. a tax-exempt counterparty). Then, the lender will receive a manufactured dividend, which is a gross coupon.[4]

This is legal in some economies. In others, the bondholder would be taxed on the theoretical coupon he or she would have received if the bond had not been repoed out. Repoing, out the bond to avoid taxation is called coupon washing.

Example

Demand for Thai bonds for both secondary trading and investment has partly been spurred by the emergence of more domestic mutual funds, which have been launching fixed-income funds. However, foreign participation in the Thai bond market is limited because of withholding taxes.

"Nobody '.I' figured out an effective way to wash-the coupon to avoid paying withholding taxes," said one investment banker in Hong Kong. Coupon washing typically involves all offshore investor selling a bond just before the coupon payment date to a domestic counterparty. Offshore entities resident ill a country having a tax treaty with the country of the bond's origin can also serve to wash coupons.

In return, the entity washing the coupon pays the offshore investor the accrued interest earned for the period before it was sold -less a small margin. Coupon washing for Thai issues is apparently widespread but is becoming more difficult, according to some sources. (IFR, Issue 1129)

Another example of this important rcpo application is from Indonesia.

Example

A new directive from Indonesia '.I' Ministry of Finance has pl/t a temporary stop to coupon washing activities players. The new directive, among other things, requires that tax be withheld on the accrued interest investors earn from their bond holdings.

Before the directive was issued a fortnight ago, taxes were withheld only from institutions that held the bond on coupon payment date. Offshore holders of Indonesian bonds got around paying the withholding tax by having the coupons washed.

Typically, coupon-washing involves an offshore institution selling and buying its bonds -just before and after the coupon payment dates - to tax-exempt institutions in Indonesia. As such, few bond holders - domestic or offshore - paid withholding taxes on bond holdings. Because the new directive requires that accrued interest 011 bonds be withheld, many domestic institutions have stopped coupon washing for international firms. (IFR Issue 1168).

The relevance of repo to taxation issues is much higher than what these readings ind_cate., , The following example shows another use of repo.

Example

In Japan there is a transaction tax on buying/selling bonds-the transfer tax. To (cut costs), repo dealers lend and borrow

Japanese Government Bonds (JGB's) and 'mark them to market every day.

The traders don't trade the bond but trade the name registration forms (NRF). NRF are,"memos" sent to Central Bank asking for ownership change. They are delivered to local custodians. The bond remains in the hands of the original owner which will be the issuer of the NRF.

JGB trading also has a no-fail rule, that is to say failure to deliver carries a very high cost and is considered taboo. (IFR, Issue 942)

Many of the standard transactions in finance have their roots in taxation strategies as these examples illustrate.

6. Synthetics Using Repos

We will now analyze repo strategies by using contractual equations that we introduced in previ-ous chapters. We show several examples. The first example deals with using repos in cash-and-', carry arbitrage, we then manipulate the resulting contractual equations to get further synthet-ics. .

6.1. A Contractual Equation

"Let Fi be the forward price observed at time I., for a Treasury bond to be delivered at a future date T, with t < T. Suppose the bond to be delivered at time T need to have a maturity of U years. Then, at time,t,we c,an (1) buy a (T - t) + U year Treasury bond, (2) repo it out-to get the necessary cash to pay for it, and (3) hold this repo position until T. At time T, cash plus the repo interest has to be returned to the repo dealer and the bond is received. The bond will have a maturity of U years. As seen above. these steps will result in exactly the same outcome as a bond forward.

We express these steps using a contractual equation. This equation provides a synthetic forward.

Forward Purchase a U – year bond to be delivered at T.	=	Buy a T + U – t year bond at t	+	Repo the bond with term T - t

"According to this, futures positions can be fully hedged by transactions shown on the right-hand side of the equation. This contractual equation can be used in several interesting applications of repo transactions. We discuss two examples.

6.2. Swaps Versus Repo

There may be some interesting connections between strips,- swaps, and repo market strategies. For example, if strips are purchased by investors who hold them until maturity, there will be fewer whole-coupon bonds. This by itself raises the probability that these bonds will trade as "special." As a result, the repo rate will on the average be lower, since the trader who is short the instrument will have to accept a "low" repo rate to get the security that is "special" to him or her.[6]

According to some traders, this may lead to an increase in the average swap spread because the availability of cheap funding makes paying fixed relatively more attractive than receiving fixed.

REPLICATION METHODS AND SYNTHESIS

Objectives

- Study of this lesson will help you to understand creation of synthesis by dynamic replication methods which are merely a generalization of the static replication methods.

The previous lessons have dealt with *static* replication of cash flows. The synthetic constructions we discussed were static in the sense that the replicating portfolio did not need any *adjustments* until the target instrument matured or expired. As time passed, the fair value of the synthetic and the value of the target instrument moved in an identical fashion.

However, static replication is not always possible in financial engineering, and replicating portfolios may need constant adjustment (rebalancing) to maintain their equivalence with the targeted instrument. This is the case for many different reasons. First of all, the implementation of static replication methods depends on the existence of other assets that permit the use of what we called *contractual equations.* To replicate the targeted security, we need a minimum number of "right-hand-side" instruments in the contractual equation. If markets in the component instruments do not exist, contractual equations cannot be used directly and the synthetics cannot be created.

Second, the instruments themselves may exist, but they may not be *liquid.* If the components of a theoretical synthetic do not trade actively, the synthetic may not really replicate the original asset satisfactorily, even though sensitivity factors with respect to the underlying risk factors are the same. For example, if constituent assets are illiquid, the price of the original asset cannot be obtained by "adding" the prices of the instruments that constitute the synthetic. These prices cannot be readily obtained from markets. Replication can only be done using assets that are liquid and "similar" but *not* identical to the components of the synthetic. Such replicating portfolios may need periodic adjustments.

Third, the asset to be replicated can be highly *nonlinear.* Using linear instruments to replicate, nonlinear assets will involve various approximations. At a minimum, the replicating portfolio_ need to be rebalanced periodically. This would be the case with assets containing optionally. As the next two chapters will show, options are convex instruments, and their replication requires dynamic hedging and constant rebalancing.

Finally, the parameters that playa role in the valuation of an asset may change, and this may require rebalancing of the replicating portfolio.

In this lesson, we will see that creating synthetics by *dynamic* replication methods follows the same general principles as those used in static replication, except for the need to rebalance periodically. In this sense, dynamic replication may be regarded as merely a generalization of the static replication methods discussed earlier. In fact, we could have started the book with principles of dynamic replication and then shown that, under some special conditions, one would end up with static replication. Yet, most "bread-and-butter" market techniques are based on the static replication of basic instruments. Static replication is easier to understand, since it is less complex. Hence, we dealt with static replication methods first. This chapter extends them now to dynamic replication.

2. An Example

Dynamic replication is traditionally discussed within a theoretical framework. It works "exactly" only in continuous time, where continuous, infinitesimal rebalancing of the replicating portfolio is possible. This exactness in replication may quickly disappear with transaction costs, jumps, in asset prices, and other complications. In discrete time, dynamic replication can be regarded: as an approximation. Yet, even when it does not lead to the exact replication of assets, dynamic replication is an essential tool for the financial engineer.

In spite of the many practical problems, discrete time *dynamic hedging* forms the basis of pricing and hedging of many important instruments in practice. The following reading shows how dynamic replication methods are spreading to areas quite far from their original use in financial engineering - namely, for pricing and hedging plain vanilla options.

Example

A San Francisco-based institutional asset manager is selling an investment strategy that uses synthetic bond options to supply a guaranteed minimum return to investors. . .

Though not a new concept – option replication has been around since the late 1980's … the bond option replication portfolio…replicates call options in that it allows investors to participate in unlimited upside while not participating in the downside.

The replicating portfolio mimics the price behaviour of the option every day until expi-ration. Each day the model provides a hedge ratio or delta, which shows how much the option price will change as the underlying asset changes.

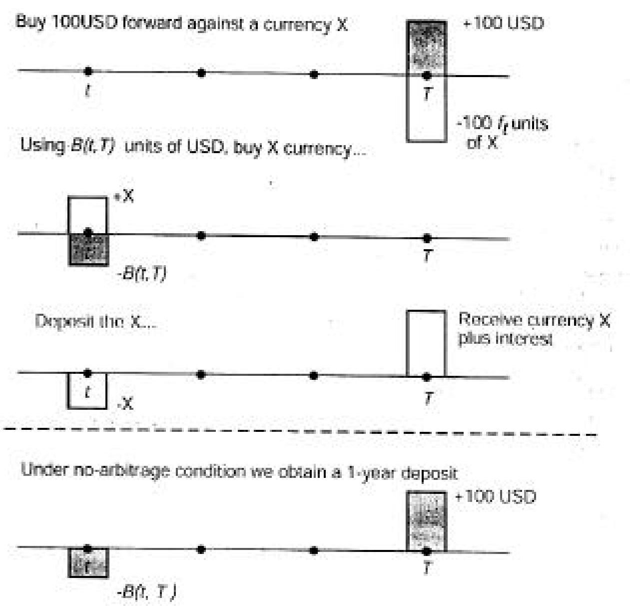
"They are definitely taking a dealer's approach, rather than an asset manager's approach in that they are not buying options from the Street; they are creating them themselves," [a dealer] said (IFR, February 28, 1998)

This reading illustrates *one* use of dynamic replication methods. It shows that market participants may replicate nonlinear assets in a cheaper way than buying the same security from the dealers. In the example, dynamic replication is combined with *principal preservation* to obtain, a product that investors may find more attractive. Hence, dynamic replication is used to create synthetic options that are more expensive in the marketplace.

3. A Review of Static Replication

In the following, we are briefly reviewing the steps taken in static replication.

1. First, we write down the cash flows generated by the asset to be replicated. Figure 7-1 repeats the example of replicating a deposit. The figure represents the cash flows of a *T*-maturity *Eurodeposit.* The instrument involves two cash flows at two different times, *t* and *T,* in a given currency, U.S. dollars (USD).
2. Next, we decompose these cash flows in order to recreate some (liquid) assets such that a vertical addition of the new cash flows matches those of the targeted asset. This is shown in the top part of Figure 7-1. A forward currency contract written against a currency *X*, a foreign deposit in currency X, and a spot FX operation have cash flows that duplicate the cash flows of the Eurodeposit when added vertically.
3. Finally, we have to make sure that the (credit) risks of the targeted asset and the proposed, synthetic are indeed the same. The constituents of the synthetic asset form, what we call the replicating portfolio.



We have seen several examples to creating such synthetic assets. It is useful to summarize two important characteristics of these synthetics.

First of all, a synthetic is created at time *t* by taking positions on three *other* instruments. But, and this is the point that we would like to emphasize, once these positions are taken we *never* again have to modify or readjust the *quantity* of the instruments purchased or sold until the expiration of the targeted instrument. This is in spite of the fact that, market risks would certainly change during the interval *(t,T).* The decision concerning the weights of replicating portfolio is made at time t, and it is kept until time *T.* As a result, the synthetic does not require further *cash injections* or *cash withdrawals,* and it matches all the cash flows generated by the original instrument.

Second, the goal is to match the expiration cash flows of the target instrument. Because the replication does not require any cash injections or withdrawals during the interval *[t, T],* the time t value of the target instrument will then match the value of the synthetic.

3.1 The Framework

Let us show how nonexistence or illiquidity of markets and the convexity of some instruments change the methodology of static synthetic asset creation. We first need to illustrate the difficulties of using static methods under these circumstances. Second, we need to motivate *dynamic* synthetic asset creation.

The treatment will naturally be more technical than the simple approach adopted prior to this chapter. It is clear that, as soon as we move into the realm of portfolio rebalancing and dynamic replication, we will need a more analytical underlying framework. In particular, we need to be more careful about the timing of adjustments, and especially how they can be made *without* any cash injections or withdrawals.

We adopt a simple environment of dynamic synthetic asset creation using a basic example- we use discount bonds and assume that risk-free borrowing and lending is the only other asset that exists. We assume that there are no markets in FX, interest rate forwards, and Eurodeposit accounts beyond the very short maturity. We will try to create synthetics for discount bonds in this simple environment. Later in the chapter, we move into equity instruments and options, and show how the same techniques can be implemented there.

We consider a sequence of intervals of length d:

$$t_0 < \ldots < t_i < \ldots < T \tag{1}$$

with

$$t_{i+1} - t_i = d \tag{2}$$

Suppose the market practitioner faces only *two* liquid markets. The first is the market for one period lending borrowing, denoted by the symbol B_t^1. The B_t is the time *t* value of $1 invested at time t_0. Growing at the annual floating interest rate L_{ti} with tenor ä, the value of B_t at time, t_n can be expressed as

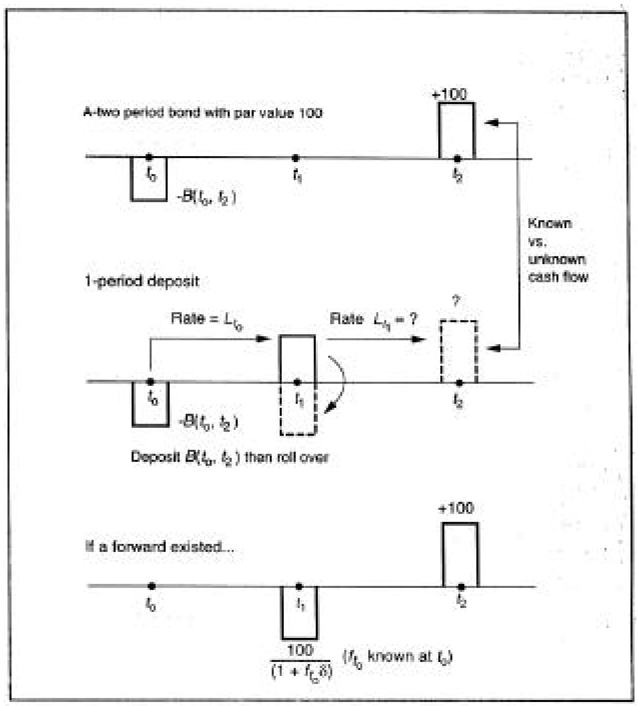
$$Bt_n = (1 + L_{t0} d)(1 + L_{t1} d)\ldots(1 + L_{tn-1} d) \tag{3}$$

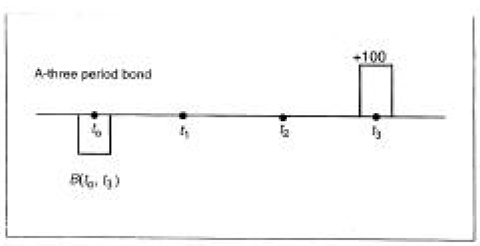
The second liquid market is for a default-free pure discount bond whose time-t price is denoted by *B(t,T).* The bond pays 100 at time T and sells for the price *B(t, T)* at time t. The practitioner can use only these two liquid instruments, *{Bt, B(t, T)},* to construct synthetics. No other liquid instrument is available for this purpose.

It is clear that these are not very realistic assumptions except maybe for some emerging markets where there is a liquid overnight borrowing-lending facility and one other liquid, on-the-run discount bond. In mature markets, not only is there a whole set of maturities for borrowing and lending and for the discount bond, but rich interest rate and FX derivative markets also exist. These facilitate the construction of complex synthetics as seen in earlier chapters. However, for discussing dynamic synthetic asset creation, the simple framework selected here will be very useful. Once the methodology is understood, it will be straightforward to add new markets and instruments to the picture.

3.2. Synthetics with a Missing Asset

Consider a practitioner operating in the environment just described. Suppose this practitioner would like to *buy*, at time *to*, a two-period default-free pure discount bond denoted by $B(t_o, T_2)$ with maturity date $T_2 = t_2$. It turns out that the only bond that is liquid is a *three-period* bond with price *B(to, T3)* and maturity $T_3 = t_3$. The $B(t_o, T_2)$ either does not exist or is illiquid. Its current fair price is unknown. So, the market practitioner decides to create the *B(to, T2)* synthetically.





One immediate consideration is that a *static* replication would *not* work in this setting. To see this, consider Figures 7-2 and 7-3. Figure 7-2 shows the cash flow diagrams for B_t. the one-period borrowing/lending, combined with the cash flows of a two-period bond.

The top portion of the figure shows that $B(t_o, T_2)$ is paid at time t_o to buy the bond that yields 100 at maturity T_2. These simple cash flows cannot, unfortunately, be reconstructed using one-period borrowing/lending B_t *only*, as can be seen in the second part of Figure 7-2. The two- period bond consists of two known cash flows at times to and *T2*. It is impossible to duplicate, *at time* t_o, the cash flow of 100 at T_2 using *Bt*, *without* making any cash injections and withdrawals, as the next section will show.

3.2.1. A Synthetic That Uses B_t Only

Suppose we adopt a rollover strategy: (1) lend money at time to for one period, at the known rate L_{to}, (2) collect the proceeds from this at it, and (3) lend it again at time t_1 at a rate L_{t1}, so as to achieve a net cash inflow of 100 at time t_2. There are *two* problems with this approach. First, the rate L_{t1} is n*ot known* at time to, and hence we cannot decide, at to, how much to lend in order to duplicate the time t_2 cash flow. The amount

$$\frac{100}{(1 + L_{t0}d)(1 + L_{t1}d)} \quad (4)$$

that needs to be invested to recover the l00 USD needed at time t_2 is not known. This is in spite of the fact that L_{to} *is* known.

Of course, we could guess how much to invest and then make any necessary additional cash injections into the portfolio when time t_1 comes: We can invest B_{to} at t_o, and then once L_t, is observed at t_1, we add or subtract an amount f1B of *cash* to make sure that

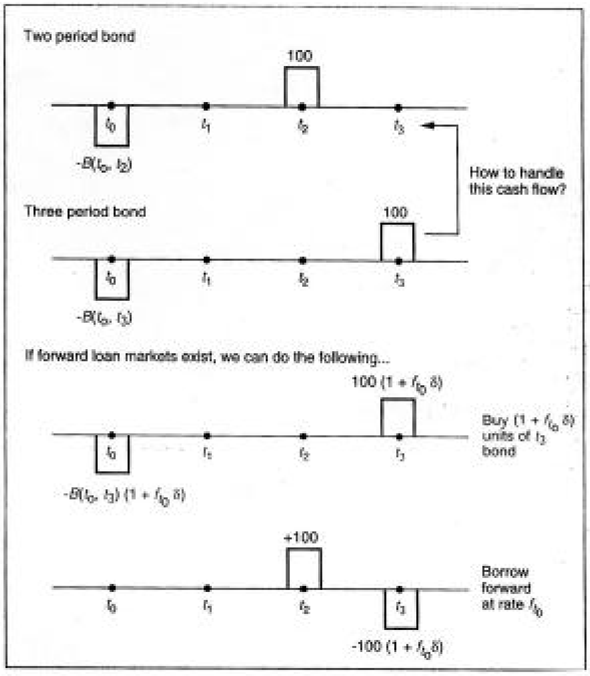
$$[B_{to}(1 + L_{to}d) + ?B](1 + L_{t1}d) = 100 \quad (5)$$

But, and this is the second problem, this strategy requires *injections* or *withdrawals* "*B* of an unknown amount at t_1. This makes our strategy useless for hedging, as the portfolio is not self-financing and the need for additional funds is not eliminated.

Pricing will be imperfect with this method. Potential injections or withdrawals of cash imply that the *true* cost of the synthetic at time to is not known.[2] Hence, the one-period Borrowing/lending cannot be used by itself to obtain a static synthetic for *B(to, T2)*. As of time t_0, the creation of the synthetic is not complete, and we need to make an additional decision at- date t_1 to make sure that the underlying cash flows match those of the targeted instrument.

3.2.2. Synthetics That Use B_t and $B(t, T_3)$

Bringing in the liquid longer-term bond *B(t,* 1:1) will not help in the creation of a *static* synthetic either. Figure 7-4 shows that- no matter what we do at time t_0 the three-period bond will have



an *extra* and nonrandom cash now of \$100 at maturity date T_3 . This cash now, being “extra,” an exact duplication of the cash flows generated by *B(t, T2)* as of time to, is not realized.

Up to this point, we did not mention the use of interest rate forward contracts. It is clear, that a straightforward synthetic for $B(t_0, T_2)$ could be created if a market for forward loans or forward rate agreements (FRAs) existed along with the “long” bond $B(t_0, T_3)$. In our particular case, a 2 x 3 FRA would be convenient as shown in Figure 7-4. The synthetic consists of buying (l + f_{to}*ä)* units of the (B_{to}, T_3) and, at the same time, taking out a one-period forward loan at the forward rate ft_0 This way, we would successfully recreate .the two-period bond in a *static* setting. But this approach assumes that the forward markets exist and that they are liquid. If these markets do not exist, dynamic replication is our only recourse.

4. “Ad..hoc” Synthetics

Then how can we replicate the two-period bond? There are several answers to this question, depending on the level of accuracy a financial engineer expects from the “synthetic.” An accurate synthetic requires dynamic replication which will be discussed later in this chapter. But, there are also less accurate, ad-hoc, solutions. As an example, we consider a simple, yet quite popular way of creating synthetic instruments in the fixed-income sector, referred to as the *immunization* strategy.

. In this section we will temporarily deviate from the notation used in the previous section and let, for simplicity, *ä* = 1. So that the t_i represents years. We assume that there are three instruments. They depend on the *same* risk factors, yet they have different *sensitivities* due to strong nonlinearities in their respective valuation formulas. We adopt a slightly more abstract framework compared to the previous section and let the three assets $\{S_{1t}, S_{2t}, S_{3t}\}$ be defined by the pricing functions:

$$S_{1t} = f(x_t) \qquad (6)$$

$$S_{2t} = f(x_t) \qquad (7)$$

$$S_{3t} = f(x_t) \qquad (8)$$

where the functions *h(.}, f(.},* and *g(.}* are nonlinear. The x_t is the common risk factor to all prices. The S_{1t} will play the role of targeted instrument, and the $\{S_{2t}, S_{3t}\}$ will be used to form the synthetic.

We again begin with *static* strategies. It is clear that as the sensitivities are different, a static methodology such as the one used in Chapters 3 through 6 cannot be implemented. As time passes, *Xt* will change randomly, and the response of S_{it}, i = 1,2,3 to changes in x_t will be different. However, one ad-hoc way of creating a synthetic for S_{1t} by using S_{2t} and S_{3t} is the following.

At time t we form a portfolio with a value equal to S_{1t}. and with weights $?_2$ and $?^3$such that the sensitivities of the portfolio

$$?^2 S_{2t} + ?^3 S_{3t} \qquad (9)$$

with respect to the risk factor x_t are as close as possible to the corresponding sensitivities of S_{1t}

Using the first-order sensitivities, we obtain two equations in two unknowns, $\{?^2, ?^3\}$

$$S = ?^2 + ?^2 S_3 \qquad (10)$$

$$\frac{\partial S_1}{\partial x} = ?^2 \frac{\partial S_2}{\partial x} + ?^3 \frac{\partial S_3}{\partial x}$$

A strategy using such a system may have some important shortcomings. It will in general-require cash injections or withdrawals over time, and this violates one of the requirements of a synthetic instrument. Yet, under some circumstances, it may provide a practical solution to problems faced by the financial engineer. The following section presents an example.

4.1. Immunization

Suppose that, at time *to,* a bank is considering the purchase of the previously mentioned two -period discount bond at a price $B(t_o, T_2)$, $T_2 = t_o + 2$. The bank can fund this transaction either by using 6-month floating funds or by selling short a three-period discount bond $B(t_o, T_3)$, $T_3 = t_o + T_3$ or a combination of both. How should the bank proceed?

The issue is similar to the one that we pursued earlier in this chapter-namely, how to construct a synthetic for $B(t_o, T_2)$. The best way of doing this is, of course, to determine an exact synthetic that is liquid and least expensive- using the 6-month funds and the three-period bond - and then, if a hedge is desired, *sell* the synthetic. This will also provide the necessary funds for buying $B(t_o, T_2)$. The result will be a fully hedged position where the bank realizes the bid-ask spread. We will learn later in the chapter how to implement this “exact” approach using dynamic strategies.

An approximate way of proceeding is to *match the sensitivities* as described earlier. In particular, we would try to match the *first-order* sensitivities of the targeted instrument. The following strategy is an example for the *immunization* of a fixed-income portfolio. In order to work with a simple risk factor, we assume that the yield curve displays parallel shifts only. This assumption rarely holds, but it is still used quite frequently by some market participants as a first-order approximation. In our case, we use it to simplify the exposition.

Example

Suppose the zero-coupon yield curve is flat at y = 8% and that the shifts are parallel. Then, the values of the 2-year, 3-year and 6-month bonds in terms of the corresponding yield y will be given by

$$B(t_0, T_2) = \frac{100}{(1+y)^2} = 85.73 \tag{12}$$

$$B(t_0, T_3) = \frac{100}{(1+y)^3} = 79.38 \tag{13}$$

$$B(t_0, T_5) = \frac{100}{(1+y)^{0.5}} = 96.23 \tag{14}$$

Using the "long" bond $B(t_o, T_3)$ and the "short" $B(t_o, T_5)$, we need to form a port-folio with initial cost 85.73. This will equal the time-to value of the target instrument.

$B(t_o, T_2)$ We also want the sensitivities of this portfolio with respect to y to be the same as the sensitivity of the original instrument. We therefore need to solve the equations

$$?^1 B(t_o, T_3) + ?^2 B(t_o, T_5) = 85.73 \tag{15}$$

$$\theta^1 \frac{\partial B(t_0, T_3)}{\partial y} + \theta^2 \frac{\partial B(t_0, T_3)}{\partial y} = \frac{\partial B(t_0, T_2)}{\partial y} \tag{16}$$

We can calculate the "current" values of the partials:

$$\frac{\partial B(t_0, T_5)}{\partial y} = \frac{-50}{(1+y)^{1.5}} = -44.55 \tag{17}$$

$$\frac{\partial B(t_0, T_2)}{\partial y} = -158.77 \tag{18}$$

$$\frac{\partial B(t_0, T_2)}{\partial y} = -220.51 \tag{19}$$

Replacing these in (15) - (16) *we get*

$$?^1 79.38 + ?^2 96.23 = 85.73 \tag{20}$$

$$?^1 (220.51) + ?^2 (44.55) = 158.77 \tag{21}$$

Solving

$$?^1 = 0.65, \ ?^2 = 0.36 \tag{22}$$

Hence, we need to short 0.65 units of the 6-month bond and short 0.36 units of the 3-year bond to create an approximate synthetic that will fund the 2-year bond. This will generate the needed cash and has the same first-order sensitivities with respect to changes in y at time to. This is a simple example of immunizing a fixed-income portfolio.

According to this, the asset being held, $B(t_o. T_2)$, is "funded" by a portfolio of other assets, in a way to make the response of the total position insensitive to first-order changes in y. /n this sense, the position is "immunized."

The preceding example shows an approximate way of obtaining "synthetics" using dynamic methods. In our case, portfolio weights were selected so that the response to a small change in the yield, *dy*, was the same. But, note the following important point.

The second and higher-order sensitivities were not matched. Thus, the funding portfolio was not really an exact synthetic for the original bono $B(t_0, T2)$. In fact, the second partials of the "synthetic" and the target instrument would respond differently to *dy*. ' Therefore, the portfolio weights $?^i$,*i*= 1,2 need to be recalculated as time passes and " new values of *y* are observed.

It is important to realize in what sense(s) the method is approximate. Even though we can adjust the weights $?^i$ as time passes, these adjustments would normally *require cash injections or withdrawals.* This means that the portfolio is not self-financing. In addition, the shifts in the yield curve are rarely parallel, and the yields for the three instruments may change by different amounts, destroying the equivalence of the first-order, sensitivities as well.

5. Principles of Dynamic Replication

We now go back to the issue of creating a satisfactory synthetic for a "short" bond $B(t_o, T_2)$ using the savings account B_t and a "long" bond $B(t_o, T_3)$' The best strategy for constructing a synthetic for $B(t_o, T_2)$ consists of a "clever" position taken in B_t and $B(t_o, T_3)$ such that, at time t_1, the extra cash *generated* by the B_t adjustment is sufficient for adjusting the $B(t_o, T_3)$.

In other words, we give up *static* replication, and we decide to adjust the time-to positions, at time t_0 in order to match the time T_2 cash payoff of the two-period bond. *However,* we adjust the positions in a way that *no net cash injections or withdrawals take place.* Whatever cash is needed at time t_1 for the adjustment of one instrument, will be provided by the adjustment of the *other* instrument. If this is done while at the same time it is ensured that the time T_2 value of this adjusted portfolio is 100, replication will be complete. It will not be static; it will require adjustments, but, importantly, we *would* know, at time t_o, how much cash to put down in order to receive $100 at T_2. Such a strategy works because both B_{tl} and $B(t_o, T_3)$ depend on the same L_{t1} the interest rate that is unknown at time *to,* and both have known valuation formulas. By cleverly taking offsetting positions in the two assets we may be able to eliminate the effects of the unknown L_{t1} as of time t_o.

The strategy will combine imperfect instruments that are correlated with each other to get a synthetic at time t_o. However, this synthetic will need constant rebalancing due to the dependence of the portfolio weights on random variables unknown as of time *to.* Yet, if these random variables were correlated in a certain fashion, these correlations can be used *against each other* to eliminate the need for cash injections or withdrawals. The cost of forming the portfolio at *to* would then equal the arbitrage-free value of the original asset.

What are the general *principles* of dynamic replication according to the discussion thus far?

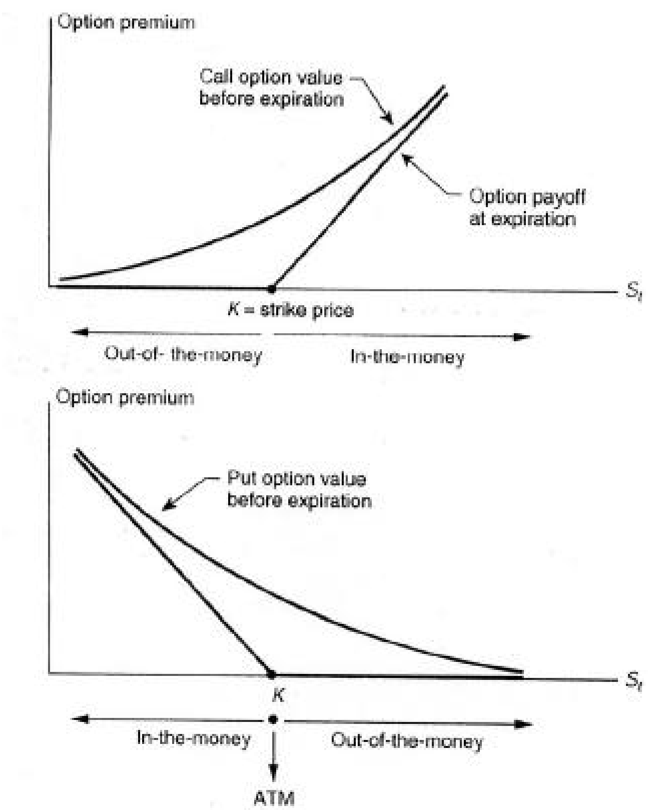
1. We need to make sure that during the life of the security there are no dividends, or other payouts. The replicating portfolio must match the *final* cash flows exactly.
2. During the replication process, there should be no net cash injections or withdrawals. The cash deposited at the initial period should equal the *true cost* of the strategy.

3. The credit risks of the proposed synthetic and the target instrument should be the same.

As long as these principles are satisfied, any replicating portfolio whose weights change, during (t,T) can be used as a synthetic of the original asset. In the rest of the chapter we apply, these principles to a particular selling and learn the mechanics of dynamic replication.

5.1. Dynamic Replication of Options

For replicating options, we use the same logic as in the case of the two-period bond discussed in the previous section. We will explore options in the next chapter. However, for completeness we repeat a brief definition. A European call option entitles the holder to buy an underlying asset, S_t, at a strike price K, at an expiration date T. Thus, at time T, $t < T$, the call option payoff is given by the broken line shown in Figure 7-5. If price at time T is lower than K, there is no payoff. If S_T exceeds K, the option is worth $(S_T - K)$. The value of the option *before* expiration involves an additional component called the *time value* and is given by the curve shown in Figure 7-5.



Let the underlying asset be a stock whose price is St: Then, when the stock price rises, the option price also rises everything else being the same. Hence the stock is highly *correlated* with the option.

This means that we can form at time *to* a portfolio using B_{t0} and S_{t0} such that as time passes, the gains from adjusting one asset compensate the losses from adjusting the other. Constant rebalancing can be done without cash injections and withdrawals, and the final value of the portfolio would equal the expiration value of the option. If this can be done with reasonably close approximation, the cost of forming the portfolio would equal the arbitrage-free value of the option. We will discuss this case in full detail later in this chapter. We will see an example when interest rates are assumed to be constant.

5.2. Dynamic Replication in Discrete Time

In practice, dynamic replication cannot be implemented in continuous time. We do need some time to adjust the portfolio weights, and this implies that dynamic strategies need to be analyzed in discrete time. We prefer to start with bonds again, and then move to options. Suppose we want to replicate the two-period default-free discount bond $B(t_o, T_2)$, $T_2 = t_2$, using B_t, $B(t_o, T_3)$ with $T2 < T3$ similar to the special case discussed earlier. How do we go about doing this in practice?

5.2.1. The Method

The replication period is $[t_o, T_2]$ and rebalancing is done in discrete intervals during this period. First we select an interval of length “ and divide the period $[t_o, T_2]$ into n such finite intervals:

$$n\Delta = T_2 - t_0 \tag{23}$$

At each $t_i = t_i\text{-}l + \Delta$, we select new portfolio weights $?_{ti}$ such that

1. At T_2, the dynamically created synthetic has exactly the same value as the T_2 - *maturity* bond.
2. At each step, the adjustment of the replicating portfolio requires no net cash injections or withdrawals.

To implement such a replication strategy, we need to deviate from static replication methods and make some *new* assumptions. In particular we just saw that correlations between the underlying assets play a crucial role in dynamic replication. Hence, we need a *model* for the way B_t, $B(t, T_2)$, and $B(t, T_3)$ move *jointly* over time.

This is a delicate process, and there are at least three approaches that can be used to model these dynamics: (1) binomial-tree or trinomial-tree methods; (2) partial differential equation (PDE) methods, which are similar to trinomial-tree models but are more general; and (3) direct modeling of the risk factors using stochastic differential equations and Monte Carlo simulation. In this section we select the *simplest* binomial-tree methods to illustrate important aspects of creating synthetic assets dynamically.

5.3. Binomial Trees

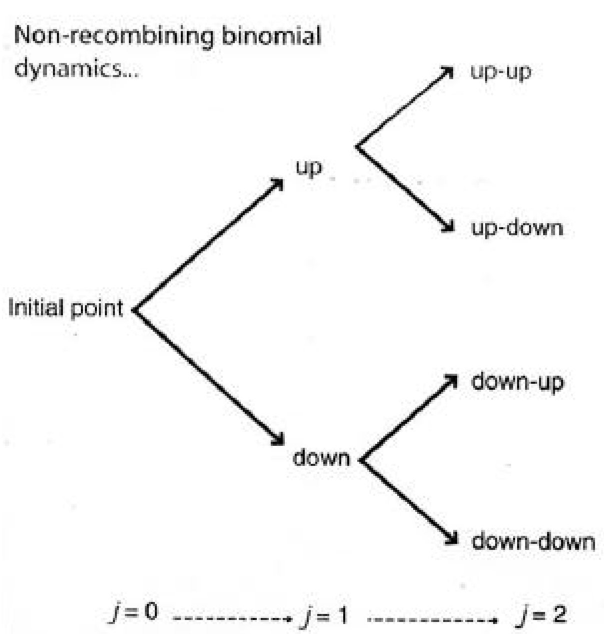
We simplify the notation significantly. We let j = 0, 1, 2 . . . denote the “time period” for the binomial tree. We chose “. so that n = 3. The tree will consist of three periods. j = 0, 1 and 2. At each *node* there are two possible states only. This implies that at j = 1 there will be two possible states and at j = 2 there will be four altogether.[3]

In fact by adjusting the “and selecting the number of possible states at each node as two three or more we obtain more and more complicated trees. With two possible states at every, node, the tree is called binomial; with three possible states, the tree is called trinomial: The implied binomial tree is in Figure 7-6. Here, possible states at every node are denoted, as usual by *up* or *down.* These terms do not mean that a variable necessarily goes up or down. They are just. shortcut names used to represent what traders may regard as "bullish" and "bearish" movements.

5.4. The Replication Process

In this section, we let " = 1, for notational convenience. Consider the two binomial trees shown in Figure 7-7 that give the joint dynamics of B_t and $B(t, T)$ over time, The top portion of the figure represents a binomial tree that describes an investment of $1 at j = 0. This investment, called *the savings account,* is rolled over at the going spot interest rate. The bottom part of the figure describes the price of the "long bond" over time. The initial point j = 0 is equivalent to t_o, and j = 3 is equivalent to t_3 when the *long bond* $B(t_o, T_3)$ matures. The tree is *l1onrecombining,* implying that a fall in interest rates following an increase would not give the same value as an increase that follows a drop. Thus, the *path* along which we get to a time node is important.[4]

We now consider the dynamics implied by these binomial trees.



5.4.1. The B_t, $B(t, T_3)$ Dynamics

First consider a tree for the B_t the savings account or risk-free borrowing and lending. The practitioner starts at time to with one dollar. The observed interest rate at j = 0 is 10%, and the dollar invested initially, yields 1.10 regardless of which state of the world is realized at time j = 1[5] There are two possibilities at j = 1. The up state is an environment where interest rates *have fallen* and bond prices, in general, have increased. Figure 7-7 shows a new spot rate *of* 8% for the up state in period j = 1. For the *down* state, it displays a spot rate that has increased to 15%.

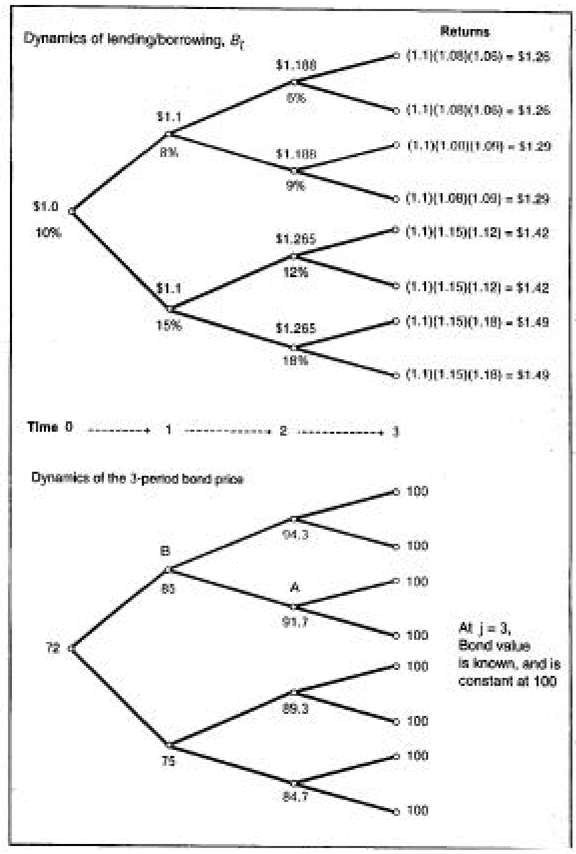
Thus, looking at the tree from the initial point to, we can see four possible paths for the spot rate until maturity time t_2 of the bond under consideration. Starting from the top, the spot interest rate paths are

{10%,8%,6%} (24)

{10%, 8%, 9%} (25)

{10%, 15%, 12%} (26)

{10%, 15%, 18%} (27)



These imply four possible paths for the value of savings account B_t

{1,1.10, 1.188, 1.26} (28)

{1, 1.10, 1.188, 1.29} (29)

{1, 1.10, 1.26, 1.42} (30)

{1, 1.10, 1.26, 1.49} (31)

It is clear that as the " becomes smaller, and the *n* gets larger, the number of possible paths will increase.

The tree for the *"long" bond* is shown in the bottom part of Figure 7-7. Here the value of the bond is $100 at j = 3, since the bond matures at that point. Because there is no default risk, the maturity value of the bond is the same regardless of which state of the world occurs. This means that one period *before* maturity, the bond will mimic a one-period risk-free investment. In fact, no matter which one of the next two states occurs, in going from a node at time j = 2 to a relevant node at time j = 3, we always invest a constant amount and receive 100. For example, at point *A*. we pay

$B(2,3)^{down}$ = 91.7 (32)

for the bond and receive 100, regardless of the spot rate move. This will change, however, as we move toward the origin. For example, at point *B*, we have either a "good" return:

$$R^{up} = \frac{94.3}{85.0} \qquad (33)$$

or a "bad" relurn:

$$R^{down} = \frac{91.7}{85.0} \qquad (34)$$

5.4.2. Mechanics of Replication

Hence, Figure 7-7 shows the dynamics of two different default-free fixed-income instruments: the savings account *Bt*, which can also be interpreted as a shorter maturity bond, and a three-period long bond $B(t, T_3)$. The question is how to combine these two instruments so as to form a synthetic medium-term bond $B(t, T_2)$.

We will now discuss the mechanics of replication. Consider Figure 7-8, which represents a binomial tree for the price of a two-period bond, $B(t, T_2)$. This tree is assumed to describe exactly the same states of the world as the ones shown in Figure 7-7. The periods beyond j = 2 are not displayed. given that the $B(t, T_2)$ matures then. According to this tree, we know the value of the two-period bond only at *j* = 2. This value is 100, since the bond matures. Earlier values of the bond are not known and hence, are left blank. The most important unknown is, of course, the time j = 0 value $B(t_o, T_2)$. This is the "current" price of the two-period bond. The problem we deal with in this section is how to "fill in" this tree

The idea is to use the information given in Figure 7-7 to form a portfolio with (time-varying) weights $?_t^{lend}$ and $?_t^{bond}$ for B_t and $B(t, T_3)$. The portfolio should mimic the value of the medium-term bond $B(t, T_2)$ at *all* nodes at j = 0,1,2. The first condition on this portfolio is that, at T_2, its value must equal 100.

The second important condition to be satisfied by the portfolio weights is that the j = 0, 1 adjustments do not require any cash injections or withdrawals. This means that, as the portfolio weights are adjusted or *rebalanced*, any cash needed for increasing the weight of *one* asset should

come from adjustment of the *other* asset. This way, cash flows will consist of a payment at time t_o, and a receipt of $100 at time T_2, with no interim net payments or receipts in between-which is exactly the cash flows of a two-period discount bond.

Then, by arbitrage arguments the value of this portfolio should track the value of the *B(t, T2)* at all relevant times. This means that the $?_t^{lend}$ and $?_t^{bond}$ will also satisfy

$$?^{lend} B_t + ?_t^{bond} B(t, T_3) = B(t, T_2) \qquad (35)$$

for all *t*, or *j.*

5.4.3. Guaranteeing Self, Financing

How can we guarantee that the adjustments of the weights $?_j^{lend}$ and $?_j^{bond}$ observed along the tree paths, j = 0,1,2 will not lead to any cash injections or withdrawals? The following additional conditions at j = 0, 1, will be sufficient to do this:

$$?_j^{lend} B_{j+1}^{up} + ?_j^{bond} B(j+1,3)^{up} = ?_{j+1}^{lend} B_{j+1}^{up}\, ?_{j+1}^{bond} B(j+1,3)^{up} \qquad (36)$$

$$?_j^{lend} B_{j+1}^{down} + ?_j^{bond} B(j+1,3)^{down} = ?_{j+1}^{lend} B_{j+1}^{down} + ?_{j+1}^{bond} B(j+1,3)^{down} \qquad (37)$$

Let us see what these conditions mean. On the left-hand side, the portfolio weights have the subscript j, while the asset prices are measured as of time j + 1. This means that the left-hand side is the value of a portfolio *chosen* at time j, and valued at a new *up* or *down* state at time. j + 1. The left-hand side is, thus, a function of "new" asset prices, but "old" portfolio weights.

On the right-hand side of these equations. we have "new" portfolio weights $?_{j+1}^{lend}$ $?_{j+1}^{bond}$ multiplied by the time j + 1 prices. Thus, the right-hand side represents the cost of a new, portfolio chosen at time j + I, either in the *up* or *downstate.* Putting these two together, the equations imply that, regardless of which state occurs, the previously chosen portfolio generates just enough cash to put together a new replicating portfolio.

If the $?_{j+1}^{lend}$ $?_{j+1}^{bond}$ are chosen so as to satisfy the equations (36) and(37), there will be no need to inject or withdraw any cash during portfolio rebalancing. The replicating portfolio will be *self-financing.* This is what we mean by dynamic replication. By following these steps we can form a portfolio at time j = 0 and rebalance at *zero* cost until the final cash flow of $100 is reached at time j = 2. Given that there is no credit risk, and all the final cash flows are equal, the initial cost of the replicating portfolio must equal the value of the two-period bond at J = 0:

$$?_0^{lend} B_0 + ?_{0bond} B(0,3) = B(0,2) \qquad (38)$$

Hence, dynamic replication would create a true synthetic for the two-period bond. Finally, consider rewriting equation (37) after a slight manipulation:

$$(?_j^{lend} - ?_{j+1}^{lend})\, B_{j+1}^{down} = -\,(?_j^{bond} - ?_{j+1}^{bond})\, B(j+1, 3)^{down} \qquad (39)$$

This shows that the cash obtained from adjusting one weight will be just sufficient for the cash needed for the adjustment of the second weight. Hence, there will be no need for extra cash injections or withdrawals. Note that this "works" even though the B_{j+1}^i and $B(j+1, 3)^i$ are random. The trees in Figure 7-7 implicitly assume that these random variables are *perfectly, correlated* with each other.

5.5. Two Examples

We apply these ideas to two examples. In the first, we determine the current value of the two- period default-free pure discount bond using the dynamically adjusted replicating portfolio from Figure 7-7. The second example deals with replication of options.

5.5.1. Replicating the Bond

The top part of Figure 7-7 shows the behavior of savings account B_t. The bottom part displays a tree for the two-period discount bond $B(t, T_3)$. Both of these trees are considered as given exogenously, and their arbitrage-free characteristic is not questioned at this point. The objective is to fill in the future and current values in Figure 7-8 and price the two-period bond $B(t, T_2)$ under these circumstances.

Example

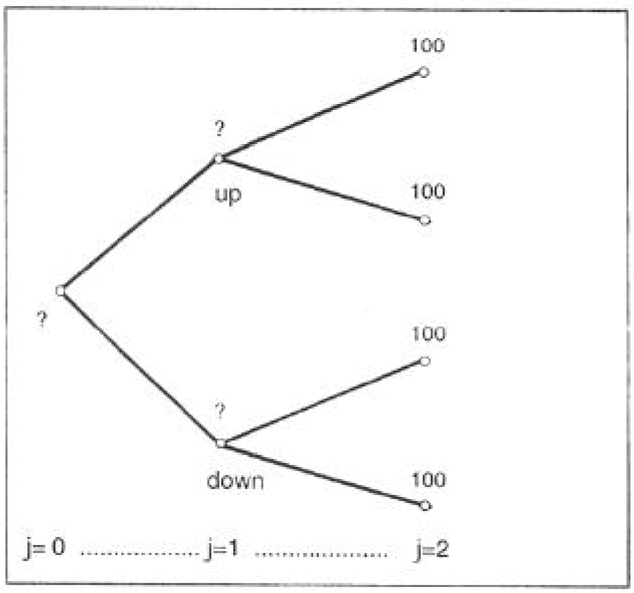
To determine the $\{B(j, 2), j = 0, 1,2\}$, *we need to begin with period* j = 2 *in Fig-ure* 7-8. *This is the maturity date for the two-period bond, and there is no default possi-bility by assumption. Thus, the possible values of the two-period bond at* j = 2, *denoted by* $B(2, 2)^i$ *,can immediately be determined:*

$$B(2,2)^{up-up} = B(2,2)^{down-up} = B(2,2)^{up-down} = B(2,2)^{down-down} = 100 \quad (40)$$

Once these are placed at the j = 2 *nodes in Figure* 7-8, *we take one step back and obtain ' the values of* $\{B(1,2)^i, i = up, down\}$. *Here, the principles that we developed earlier will be used. As "time" goes from* j = 1 *to* j = 2, *the value of the portfolio put together at* j = 1 *using* B_1 *and* $B(1,3)^i$ *should match the possible values of* B(2, 2) *at all nodes. Consider first the top node* $B(1,2)^{up}$. *The following equations need to be satisfied:*

$$?_1^{lend,up} B_2^{up-up} = ?_1^{bond,up} B(2,3)^{up-up} = B(2,2)^{up-up} \quad (41)$$

$$?_1^{lend,up} B_2^{up-down} + ?_1^{bond,up} B(2,3)^{up-down} = B(2,2)^{up-down} \quad (42)$$



Here, the ? 's have j = 1 subscript hence the left-hand side is the value of the replicating portfolio put together at time j = 1 but valued as of j = 2. In these equations, all variables are known except portfolio weights $?_1^{lend,up}$ and $?_1^{bond,up}$. Replacing from Figures 7-7

$$?_1^{lend,up} 1.888 + ?_1^{bond,up} 94.3 = 100 \quad (43)$$

$$?_1^{lend,up} 1.888 + ?_1^{bond,up} 91.7 = 100 \quad (44)$$

Solving these two equations for the two unknowns, we get the replicating portfolio weights for j = 1, i = up. The.'1e are in units of securities, not in dollars.

$$?_1^{end, up} = 84.18 \quad (45)$$

$$?_1^{bond, up} = 0 \quad (46)$$

Thus, if the market moves to i = up, 84.18 units of the B_1 will be sufficient to replicate the future values of the bond at time j = 2. In fact, this position will have the j = 2 value of

$$84.18(1.188) = 100 \quad (47)$$

Note that the weight for the long bond is zero.[6] The cost of this portfolio at time j = 1 can be obtained using the just calculated $?_1^{lend,up}$ and $?_1^{bond,up}$ this cost should equal $B(I,2)^{up}$.

$$?_1^{lend,up}(1.1) + ?_1^{bond,up}(85.0) = 92.6 \quad (48)$$

Similarly, for the state j = 1, i = down we have the two equations:

$$?_1^{lend,down} 1.265 + ?_1^{bond, down} 89.3 = 100 \quad (49)$$

$$?_1^{lend,down} 1.265 + ?_1^{bond, down} 89.3 = 100 \quad (50)$$

Solving, we get the relevant portfolio weights:

$$?_1^{lend,down} = 79.05 \quad (51)$$

$$?_2^{lend,down} = 0 \quad (52)$$

We obtain the cost of the portfolio for this state:

$$?_1^{lend,down}(1.1) + ?_1^{bond,down}(75) = 86.9 \quad (53)$$

This should equal the value of $B(1,2)^{down}$. Filially, we move to the initial period to determine the value B (0,2). The idea is again the same. At time j = 0 choose the portfolio weights $?_1^{lend}$ and $?_1^{bond}$ such that, as time passes, the value of the portfolio equals the possible future values of B(1, 2):

$$?_0^{lend} 1.1 + ?_0^{bond} 85.00 = 92.6 \quad (54)$$

$$?_0^{lend} 1.1 + ?_0^{bond} 75.00 = 86.9 \quad (55)$$

Here, the left hand side is the value of the portfolio put together at time j = 0 such that its value equals those of the two-period bond at j = 1. Solving for the unknowns,

$$?_0^{lend} = 40.1 \quad (56)$$

$$?_0^{bond} = 0.57 \quad (57)$$

Thus, at time j = 0 we need to make a deposit of 40.1 dollars and buy 0.57 units of , the three-period bond with price B(O,3). This will replicate the two possible values $\{B(1,2)^i, i = up, down\}$. The cost of this portfolio must equal the current fair value of B(0,2}, if the trees for the B_t and B(j,3) are arbitrage-free. This cost is given by

$$B(0,2) = 40.1 + 0.57(72\} = 81.14 \quad (58)$$

This is the fair value of the two-period bond at j = 0.

The arbitrage-free market value of the two-period bond is obtained by calculating all the current and future weights for a dynamic self-financing portfolio that duplicates the final cash flows of a two-period bond. At every step, the portfolio weights are adjusted so that the rebalanced portfolio keeps matching the values of *B(j,* 2}, j = 0,1,2. The fact that there were only two possible moves from every node gave a system of two equations, in two unknowns.

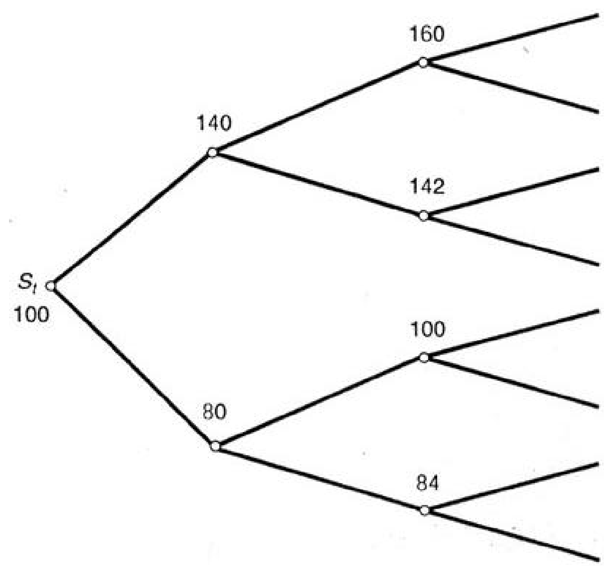
Note the (important) analogy to static replication strategies. By following this dynamic strategy and adjusting the portfolio weights, we guarantee to match the final cash flows generated by the two-period bond, while never really making any cash injections or withdrawals. Each time a future node is reached, the previously determined portfolio will always yield just enough, cash to do necessary adjustments.[7]

5.6 Application to Options

We can apply the replication technique to options, and create appropriate synthetics. Thus, consider the same risk-free lending and borrowing B_t dynamics shown in Figure 7-7. This time, we would like to replicate a call option C_t written on a stock S_t. The call has the following *plain vanilla* properties. It expires at time t_2

and has a strike price $K = 100$. The option is European and cannot be exercised before the expiration date. The underlying stock S_t does not pay any dividends. Finally, there are no transaction costs such as commissions and fees in trading S_t or C_t.

Suppose the stock price S_t follows the tree shown in Figure 7-9. Note that unlike a bond, the stock never "matures" and future values of S_t are always random. There is no terminal time period where we know the future value of the S_t, as was the case for the bond that expired at time T_3.



However, the corresponding binomial tree for the call option still has *known* values at expiration date j = 2. This is the case since, at expiration, we know the possible values that the.

option may assume due to the formula:

$$C_2 = max\ [S2 - 100, 0] \tag{59}$$

Given the values of S_2, we can determine the possible values of C_2. But, the values of the call at *earlier* time periods still need to be determined.

How can this be done? The logic is essentially the same as the one utilized in the case of two-period default-free bond. We need to determine the current value of the call option, denoted by C_0, using a dynamically adjusted portfolio that consists of the savings account and of the stock S_t.

Example

Start with the expiration period and use the boundary condition:

$$C_2^i = \max [{}_{s2}{}^i - 100,0] \tag{60}$$

where the -i subscript represents gain in the states of the world {up-up, up-down, down":" up, down-down}. Using these, we determine the four possible values of C at expiration:

$$c_2^{up\text{-}up} = 60,\ C_2^{up\text{-}down} = 42,\ C_2^{down\text{-}up} = 0,\ C_2^{down\text{-}down} = 0 \tag{61}$$

Next, we take one step back and consider the value C_1^{up}. We need to replicate this with a portfolio using B_1, S_1, such that as "time" passes, the value of this portfolio stays identical to the value of the option C_2^i. Thus, we need

$$?_1^{lend,up} B_2^{up\text{-}up} + ?_1^{stock,up} S_2^{up\text{-}up} = C_2^{up\text{-}up} \tag{62}$$

$$?_1^{lend,up} B_2^{up\text{-}down} = ?_{1stock,up} S_2^{up\text{-}down} = C_2^{up\text{-}down} \tag{63}$$

Replacing the known values from Figure 7-7 and 7-9, we have two equations and two unknowns:

$$?_1^{lend,up}\ (1.188) + ?_1^{stock,up}(160) = 60 \tag{64}$$

$$?_1^{lend,up}\ (1.188) + ?_1^{stock,up}(142) = 42 \tag{65}$$

Solving for the portfolio weights $?_1^{lend,up}$ and $?_1^{stock,up}$, we get

$$?_1^{lend,up} = -84.18 \tag{66}$$

$$?_1^{stock,up} = 1 \tag{67}$$

Thus, at time j = 1, i = up, we need to sell 84.18 units of B_t and buy one stock. The behavior of this portfolio in the immediate future will be equal to the future values of { C_2^i} where i denotes the four possible states at j = 2. The cost of this portfolio is C_1^{up}.

$$C_1^{up} = -84.18(1.1) + 140 \tag{68}$$

$$= -47.40 \tag{69}$$

Similarly, in order to determine crown, we first form a replicating portfolio by solving the equations

$$?_1^{lend,down}(1.26) + ?_1^{stock,down}(100) = 0 \tag{70}$$

$$?_1^{lend,down}(1.26) + ?_1^{stock,down}(84) = 0 \tag{71}$$

which gives

$$?_1^{lend,down} = 0 \tag{72}$$

$$?_1^{stock,down} = 0 \tag{73}$$

The cost of this portfolio is zero and hence the option is worthless if we are at j = 1, i = down:

$$C_1^{down} = 0 \tag{74}$$

Finally, the fair value Co of the option can be determined by finding the initial portfolio weights from

$$?_0^{lend}(1.1) + ?_0^{stock}(140) = 47.40 \tag{75}$$

$$?_0^{lend}(1.1) + ?_0^{stock}(80) = 0 \tag{76}$$

We obtain

$$?_0^{lend} = -57.5 \tag{77}$$

$$?_0^{stock} = .79 \tag{78}$$

Thus, we need to borrow 57.5 dollars and then buy .79 units of stock at j = 0. The cost of this will be the current value of the option:

$$C_0 = -57.5 + .79(100) \tag{79}$$

$$= 21.3 \tag{80}$$

This will be the fair value of the option if the exogenously given trees are arbitrage-free.

Note again the important characteristics of this dynamic strategy. (1) To determine the current value of the option, we started from the expiration date and used the boundary condition, (2) We kept adjusting the portfolio weights so that the replicating portfolio eventually matched the final cash flows generated by the option, (3) Finally, there were no cash injections or cash withdrawals, so that the initial amount invested in the strategy could be taken as the cost .of the synthetic.

6. Some Important Conditions

In order for these methods to work, some important assumptions were needed. Until this point we did not discuss these in detail.

6.1. Arbitrage, Free Initial Conditions

The methods discussed in this chapter will work only if we start from dynamics that originally exclude any arbitrage opportunities. Otherwise, the procedures shown will give "wrong" results. For example, some bond prices $B(j, T_2)^i$, $j = 0,1$ or the option price may turn out to be *negative.*

There are many ways the arbitrage-free nature of the original dynamics can be discussed. One obvious condition concerns the returns associated with the savings account and the other constituent asset. It is clear that, at all nodes of the binomial trees in Figure 7-7, the following condition needs to be satisfied:

$$R_j^{down} < L_j < R_j^{up}$$

where L_j is the one period spot rate that is observed at that node and the R_j^{down} and R_j^{up} are two possible returns associated with the bond at the same node.

According to this condition, the risk-free rate should be between the two possible returns that one can obtain from holding the "risky" asset B(t,T). For the case of bonds, before expiration we must also have, due to arbitrage,

$$R_j^{down} -= L_j = R_j^{up} \tag{82}$$

Otherwise, we could buy or sell the bond, and use the proceeds in the risk-free investment to make unlimited gains.

Yet, the arbitrage-free characteristic of binomial trees normally require more than this simple condition. As Chapter will show the underlying dynamics should be conformable with proper *Martingale* dynamics in order to make the trees arbitrage-free.

6.2. Role of Binomial Structure

There is also a very strong assumption behind the binomial tree structure that was used during. . the discussion. This assumption does not change the logic of the dynamic replication strategy, but, can make it numerically more complicated if it is not satisfied.

Consider Figure 7-7. In these trees, it was assumed that when the short rate dropped, the long rate always dropped along with it. Conversely, when the short rate increased, the long rate increased with it. That is to say, the long bond return and the short rate were *perfectly* correlated. It is thanks to this assumption that we were able to associate a future value of B_t with another future value of $B(t, T_3)$. These "associations" were never random. A similar assumption was made concerning the binomial trees for S_t and C_t. The movements of these two assets were perfectly correlated.

This is a rather strong assumption, and is due to the fact that we are using the so-called *one-factor* model. It is assumed that there is a single random variable that determines the future value of the assets under consideration at every node. In reality, given a possible movement in the short rate L_t, we may not know whether a bond price $B(t, T)$ will go up or down in the immediate future, since *other* random factors may be at play. Under such conditions, it would be impossible to obtain the same equations, since the *up* or *down* values of the two assets would not be associated with certainty.

Yet, introducing further random factors will only increase the numerical complexity of the tree models. We can, for example, move from binomial to trinomial or more complicated trees. The general logic of the dynamic replication does not change. However, we may need further base assets to form a proper synthetic.

7. Real Life Complications

Real-life complications make dynamic replication a much more fragile exercise than static, replication. The problems that are encountered in static replication are well known. There are operational problems, counterparty risk, and the theoretically exact synthetics may not be iden-tical to the original asset. There are liquidity problems and other transactions costs. But, all these are relatively minor and in the end, static replicating portfolios used in practice generally provide good synthetics.

With dynamic replication, these problems are magnified, because the underlying positions need to be readjusted many times. For example, the effect of transaction costs is much more serious if dynamic adjustments are required frequently. Similarly, the implications of liquidity problems will also be more serious. But more importantly, the real-life use of dynamic replication methods brings forth *new* problems that would not exist with static synthetics. We study these briefly.

7.1. Bid, Ask Spreads and Liquidity

Consider the simple case of bid-ask spreads. In static replication, the portfolio that constitutes the synthetic is put together at time t and is never altered until expiration T. In such an environment, the existence of bid-ask spreads may be nonnegligible but is hardly a major aspect of the problem. After all, any bid-ask spread will end up increasing (or lowering) the cost of the associated synthetic, and in the unlikely case that these are prohibitive then the synthetic will not be put together.

Yet, with dynamic replication, the practitioner is constantly adjusting the replicating port-folio. Such a process is much more vulnerable to widening bid-ask spreads or the underlying liquidity changes. At the time dynamic replication is initiated, the future movements of bid-ask spreads or of liquidity will not be known exactly and cannot be factored into the initial cost of the synthetic. Such movements will constitute additional risks, and increase the costs even when the synthetic is held until maturity.

7.2. Models and Jumps

Dynamic replication is never perfect in real life. It is done using *models* in discrete time. But, models imply assumptions and discrete time means approximation. This leads to a model risk. Many factors and the possibility of having jumps in the underlying risks may have serious consequences if not taken into account properly during the dynamic replication process.

7.3. Maintenance and Operational Costs

It is easy to obtain a dynamic replication strategy theoretically. But in practice, this strategy needs to be implemented using appropriate position-keeping and risk-management tools. The necessary software and human skills required for these tasks may lead to significant new costs.

7.4. Changes in Volatility

Often, dynamic replication is needed because the underlying instruments are nonlinear. It turns out that, in dealing with nonlinear instruments, we will have additional exposures to new and less transparent risks such as movements in the *volatility* of the associated risk factors. Because risk-managing volatility exposures is much more delicate (and difficult) than the management, of interest rate or exchange rate risks, dynamic replication often requires additional skills.

In the exercises at the end of this chapter we briefly come back to this point and provide a reading (and some questions) concerning the role of volatility changes during the dynamic, hedging process.

8. Conclusions

We finish the chapter with an important observation. Static replication was best done using cash flow diagrams and resulted in contractual equations with *constant* weights.

Creating synthetics dynamically requires constant adjustments and careful selection of portfolio weights $è^{i}_{t}$ in order to make the synthetic *self financing.* Thus, we again use contractual equations. But this time, the weights, placed on each contract changes as time passes. This requires the use of algebraic equations and is done with computers.

Finally, the *dynamic* synthetic is nothing but the sequence of weights $\{ ?^{i}_{1}, ?^{i}_{2}, ?^{i}_{3} \}$ that the financial engineer will determine at time t_0.

References

Several books deal with dynamic replication. Often these are intermediate-level textbooks on derivatives and financial markets. We have two preferred sources that the reader can consult for further examples. The first is jarrow (2002). This book deals with fixed-income examples only. The second is Jarrow and Turnbull (1999), where dynamic replication methods are discussed in much more detail with a broad range of applications. The reader can also consult the original Cox and Ross (1976a) article. /t remains a very good summary of the procedure.

OPTION MECHANICS

Objectives

- After completion of this lesson you will be able to identify fully of how the options market players work and at the same time to what extent you are exposed to such volatility.

You will be quite interesting in this lesson. It is an introduction to methods used in dealing with optionally in financial instruments. Compared to most existing text books, the present text adopts a different way of looking at options. We discuss options from the point of view of an options market maker. In our setting, options are not presented as instruments to bet on or hedge against the direction of an underlying risk. Instead, options are motivated as instruments of volatility.

In the traditional textbook approach, options are introduced as directional instruments. This is not how market professionals think of options. In most textbooks, a call option becomes in-the- money and hence profitable if the underlying price increases, indirectly associating it with a bullish view. The treatment of put options is similar. Puts are seen as appropriate for an investor who thinks the price of the underlying asset is going to decrease. For an end investor or retail client, such directional motivation for options may be natural. But, looking at options this way is misleading if we are concerned with the interbank or interdealer market .In fact. Motivating options as directional tools will disguise the fundamental aspect of these instruments, namely that options are tools for trading two views of options is quite different, and we would like the reader to think like an option trader or market maker.

This chapter intends to show that an option exposure, when fully put in place, is an impure position on the way volatility to increase. A market maker with a net long position in option is is someone who is "expecting" the volatility to increase. A market maker who is short the option, is someone who thinks that the volatility of the underlying is going to decrease. Sometimes such positions are taken as funding vehicles.

In this sense, a trader's way of looking at puts and calls is in complete contrast to the directional view of options. For example, market makers look at European calls and puts as if they were identical objects. As we will see in this chapter, from an option market mader's point of view, there is really no difference between buying a call or buying a put. Both of these transactions, in the end, result in the same payoff. Consider Figure 1, where we show two possible intraday trajectories of an underlying price, S_t. In one case prices are falling rapidly, while in the other prices are rising. An option trader will sell puts or calls with the same ease.

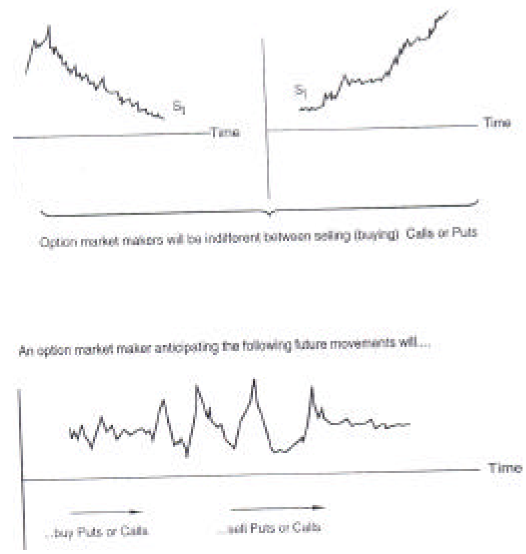


Figure 1

As we will see, rather than with the issue of which type of sell, the trader may be concerned with whither he should sell any option, or buy them.

In this lesson and the next, we intend to clarify the connection between volatility and option prices. However, we first review some basics.

1. What Is an Option?

From a market practitioner's point of view, options are instruments of volatility. A retail investor who owns a call on an asset, S_t , may feel that a persistent upward movement in the price of this asset is "good" for him or her. But, a market maker who may be long in the same call may prefer that the underlying price St oscillate as much as possible, as often as possible. The more frequently and violently price oscillate, the more frequently and violently price oscillate, the more long (short) positions in option books will gain (lose), regardless of whether calls or puts are owned.

The following reading is a good example as to how option traders look at options.

Example

Wall Street firms are gearing up to recommend long single-stock vol positions on companies about to report earnings. While earnings seasons often offer opportunities for going long vol via buying Calls or puts, this season should present plenty of opportunities to benefit from long vol positions given overall negative investor sentiment. Worse than expected earnings are released from one company and can send shockwaves through the entire market.

The big potential profit from these trades is from gamma, in other words, large moves in the underlying, rather than changes in implied vol. One promising name… announced in mid-February that manufacturing process and control issues have led to reduced sales of certain products in the U.S., which it expected to influence its first quarter and sales and earnings. On Friday, options maturing in August had a mid- market implied vol of around 43% , which implies a 2.75% move in the stock per trading day. Over the last month, the stock has been moving on average 3% a day, which means that by buying options on the company, you are getting vol cheap. (Derivatives Week, April 1, 2001)

This reading illustrates several important characteristics of options. First, we clearly see that puts and calls are considered as similar instruments by market practitioners. The issue is not to buy puts or calls, but whether or not to buy them.

Second, and this is related to the first point, notice that market participants are concerned with volatilities and not with the direction of prices – referring to volatility simply as vol. Market professionals are interested in the difference between actual daily volatilities of stock prices and the volatilities implied by the options. The last sentence in the reading is a good (but potentially misleading) example of this. The reading suggests that option imply a daily volatility of 2.75%, while the actual underlying moves more than what the option price implies on a given day.[1] This distiction between implied volatility and "actual volatility" should be kept in mind.

Finally, the reading seems to refer to two different types of gains from volatility. One, being from "large movements in the unerlying price", which leads to gamma gains, and the other from impied volatility – which leads to vega gains. During this particular episode, market professionals were expecting implied volatility to remain the same, while the underlying assets exhibited sizable fluctuations. It is difficult, at the outset, to understand this difference. The present chapter will clarify these notions and reconcile the market professional's view of options with the directional approach the reader may have been exposed to earlier.[2]

[1] This analysis should be interpreted carefully. In the option literature, these are many different measures of volatility. As this chapter will show, it is perfectly reasonable that the two values be different, and this may not necessary imply an arbitrage possibility.

[2] The previous example also illustrates a technical point concerning volatility calculations in practice. Consider the way daily volatility was calculated once annualized percentage volatility was given. Suppose there are 246 trading days in a year. Then, note that and annual percentage volatility of 43% is not divided by 246. Instead, it is divided by the square root of 246 to obtain the "daily" 2.75% volatility. This is known as the square root rule, and has to do with the role played by Wiener processes in modeling stock price dynamics. Wiener process increments have a variance that is proportional to the time that has elapsed. Hence, the standard division or volatility will be proportional to the square root of the elapsed time.

2. Options: Definition and Notation

Option contracts are generally divided into the categories of plain vanilla and exotic options, although many of the options that used to be known as exotic are vanilla instruments today. In discussing options, it is good practice to start with a simple benchmark model, understand the basics of options, and then extended the approach to more complicated instruments. This simple benchmark will be a plain vanilla option treated within the framework of Black-/Scholes model.

The buyer of an option does not buy the underlying instrument. He or she buys a right. If this right can be exercised only at the expiration date, then the option European. If it can be exercised any time during the specified period, the option is said to be American. A Bermudan option is "in – between," given that it ca be exercised at more than one of the dates during the life of the option.

In the case of a European plain vanilla call, the option holder has purchased the right to "buy" the underlying instrument at a certain price, called the strike or exercise price, at a specific date, called the expiration date. In the case of the European plain vanilla put, the option holder has again purchased the right to an action. The action in this case is to "sell" the underlying instrument at the strike price, at the expiration date.

American style options can be exercised anytime until expiration and hence may be more expensive. They may carry an early exercise premium. At the expiration date, options cease to exist. In this chapter, we discuss basic properties of options using mostly plain vanilla calls. Obviously, the treatment of puts would be similar.

2.1 Notation

We denote the strike prices by the symbol K, and the expiration date by T. The price or value of the underlying instrument will be denoted by S_t if it is a cash product, and by F_t if the underlying is a forward or futures price. The fair price of the call at time t will be denoted by C(t), and the price of the put by P(t)[3]. These prices depend on the variables and parameters underlying the contract. We use S_t as the underlying, and write the corresponding call option pricing function as

$$C(t) = C(S_t, t \mid r, K, \sigma, T) \tag{1}$$

Here, () is the volatility of S_t and r is the spot interest rate, assumed to be constant. In more compact form, this formula can be expressed as

$$C(t) = C(S_t, t) \tag{2}$$

This function is assumed to have the following partial derivatives:

$$\frac{\partial C(St,t)}{\partial St} = C_s \tag{3}$$

[3] The way we characterize and handle the time index is somewhat different than the treatment up to this chapter. Option prices are not written as C_t and P_t as the notion of previous chapters may suggest. Instead, we use the notion C(t) and P(t). The former notation will be reserved for the partial derivative of an option's price with respect to time t.

$$\frac{\partial C(St,t)}{\partial S^2_t} = C_s \quad (4)$$

$$\frac{\partial C(S_t,t)}{\partial t} = C_s \quad (5)$$

More is known on the properties of these partials. Everything else being the same. If S t is increase. The call option price, c(t), is also increase . If St declines, the price declines. But , the changes in C(Mt) will never exceed those in the underlying asset, St. Hence, We should have

$$0 < C_s < 1 \quad (6)$$

At the same time, everything else being the same , as t increases, the life of the option gets shorter and the time- value declines,

$$C_t < 0 \quad (7)$$

Finally, the expiration pay off of the call(put) option is a convex function, and we expect the c(st,t) to be convex as well. This means that

$$0 < C_{ss} \quad (8)$$

This information about the partial derivatives is assumed to be known even when the exact form of $C(S_t,t)$ itself is not known.

The notation in Equation (1) suggests that the partials themselves are function of S_t,r,K,t,T. and s .

Hence, one may envisage some further, higher order partials. The traditional Black Scholes vanilla option pricing environment uses the partials, $\{C_s,C_{ss},C_t\}$ only, Further partial derivatives are brought into the picture as the Black-Scholes assumptions are relaxed gradually.

Figure 2 shows the expiration date payoffs of plan vanilla put and call options. In the same figure we have the time t,t<T value of the calls and puts. These values trace a smooth convex curve obtained from the Black-Scholes formula.

We now consider a real- life application of these concepts, The following example looks at Microsoft options traded at the Chicago Board of Options exchange, and discusses various parameters within this context.

Example

Suppose Microsoft (MSFT) is "currently" trading at 61.15 at Nasdaq. Further the overnight rate is 2.7%. We have the following quotes from the Chicago Board of Options Exchange (CBOE).

In the table the first column gives the expiration date and strike level of the option. The exact time of expiration is the third Friday of every month. These equity options in CBOE are of American style, The bid price is the price at which the market maker is willing to buy the option from the client whereas the ask price is the price at which he or she willing to sell it to the client.

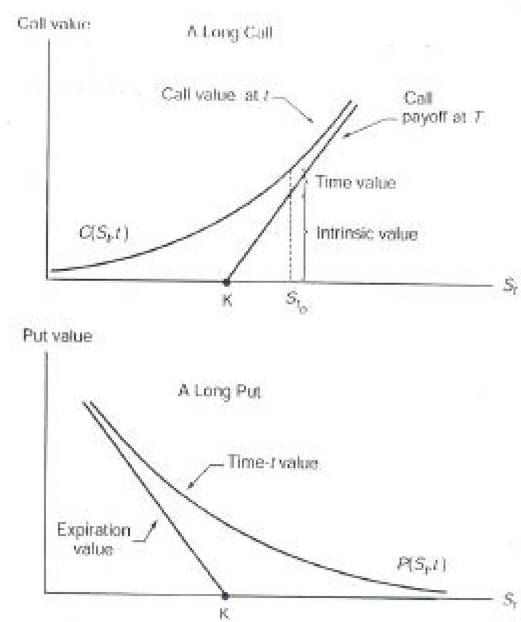


Figure .2

The call option premium is denoted by C(t). By buying the call, the client makes sure that be or she can buy one unit of the underlying at a maximum price K, the client will not exercise the option. There is no need to pay K dollars for something that is selling for less in the marketplace. The option will be exercised only if S_t equals or exceed K at time T.

Looked at this way, options are somewhat similar to standard insurance against potential increases in commodities prices. In such a framework, options can be motivated as directional instruments. One has the impression that, an increase in S_t is harmful for the client, and that the call "protects" against this risk. The situation for puts is symmetrical. Puts appear to provide protection against the risk of undesirable "declines" in S_t. In both cases, a certain direction in the change of the underlying price S_t is associated with the call or put, and these appear to be fundamentally different instruments.

Figure 3 illustrates these ideas graphically. The upper part shows the payoff diagram for a call option. Initially, at time t_0, the underlying price is at S_{to}. Note that $S_{to} < K$, and the option is out-of- the- money. Obviously, this does not mean that the right to buy the asset at time T for K dollars has no value. In fact, from a client's point of view, St may move up during interval t Î[t0, T] and end up exceeding K by time T. This will make the option in the money.

It would then be profitable to exercise the option and buy the underlying at a price K. The option payoff will be the difference S_t-K, if S_T exceeds K. This payoff can be shown either on the horizontal axis or, more explicitly, on the vertical axis.[4] Thus, looked at from the retail client's point of view, even at the price level S_{to} , the out –of – the money option is valuable, since it may become in – the – money later. Often, the directional motivation of options is based on these kinds of arguments.

If the option expires at S_T = K, the option will be at–the–money (ATM} and the option holder may or may not chose to receive the underlying. However, as the costs associated with delivery of the call underlying are, in general, less than the transactions costs of buying the underlying in the open market, some holders of ATM options prefer to exercise.

Hence, we get the typical price diagram for a plain vanilla European call option. The option price for t Î[to,T] is shown in Figure 3 as a smooth convex curve that converges to the

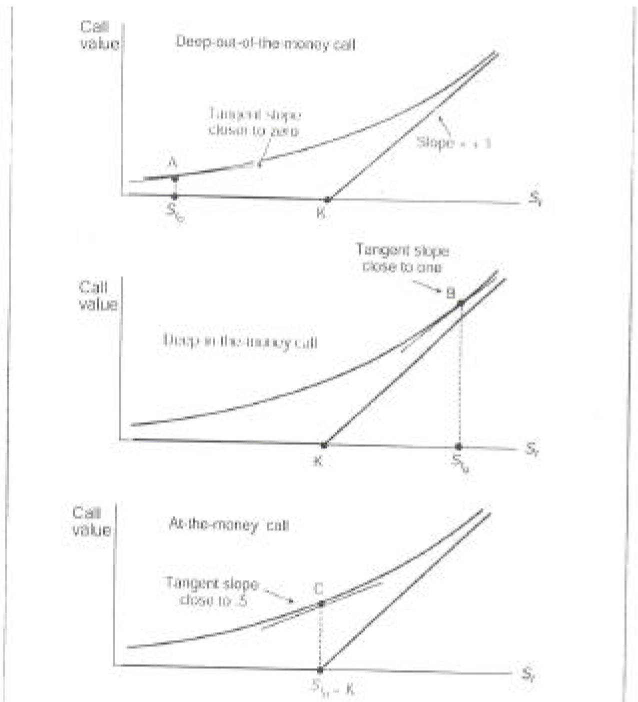


Figure 3

(As usual, the upward-sloping line in Figure 3 has sloped +1, and thus "reflects" the profit, ST – K on the horizontal axis, toward the vertical axis.)

Piecewise linear option payoff as expiration time T approaches. The vertical distance between the payoff line and the horizontal axis is called intrinsic value. The vertical distance between the option price curve and the expiration payoff is called the time value of the option. Note that for a fixed t, the time value appears to be at a maximum when the option is at-the-money - that is to say, when S_t =K.

3.3. Some Intriguing Properties of the Diagram

Consider point A in the top part of Figure 8-3. Here, at time S_t, the option is deep out –of – money. The tangent at point A has a positive slope that is little different from zero. The curve is almost "linear" and the second derivative is also close to zero. This means that for small changes in S_t , the slope of the tangent will not vary much.

Now , consider the case represented by point B in Figure 3 . Here, at time t, the option is deep in – the –money. S_t is significantly higher than the strike price. However, the time value is again close to zero. The curve approaches the payoff line and hence has a slope close to +1. Yet the second derivative of the curve is once again very close to zero. This again means that for small changes in S_t , the slope of the tangent will not vary much.

The third case is shown as point C in the lower part of Figure 3 . Suppose the option was at – the –money at time t, as shown by point C. The value of the option is entirely made of time value. Also, the slope of the tangent is close to 0.5. Finally, it is interesting that the curvature of the option is highest at the point C and that if St changes a little, the slope of the tangent will change significantly.

This brings us to an interesting point. The more convex the curve is at a point, the higher seems to be the associated time value. In the two extreme cases where the slope of the curve is diametrically different, namely at points A and B, the option has a small time value. At both points, the second derivative of the curve is small. When the curvature reaches its maximum, the time value is greatest. The question, of course, is whether or not this is a coincidence.

Pursuing this connection between time value and curvature further will lead us to valuing the underlying volatility. Suppose, by holding an option, a market maker can somehow generate "cash" earnings, as S_t oscillates. Could it be that, everything elase being the same, the greater the curvature of C(t) , the greater the cash earnings are? Our task in the next section is to show that this is indeed the case.

1. Options as Volatility Instruments

In this section we see how convexity is translated into cash earnings, as S_t oscillates and reates time value.[6] The discussion is conducted in a highly simplified environment to facilitate understanding of the relationship between volatility and cash gains (losses) of long (short) option positions.

Consider a market maker who quotes two –way prices for a European vanilla call option C(t) with strike K, and expiration T, written on a non dividend paying asset, denoted by S_t. [7] Let the

[5] That is, it will stay close to 1.

[6] It is important to emphasize that this way of considering options is from an inter-bank point of view. For end investors, options can still be interpreted as directional investments, but the pricing and hedging of options can only be understood when looked at from the dealer's point of view. The next chapter will present applications related to classical uses of options.

[7]. Remember that market makers have the obligation to buy and sell at the prices they are quoting.

Risk- free interest rate r is constant. For simplicity, consider an at – the- money option, K= S_t.

In the following, we first show the initial steps taken by the market maker who buys an option..

Then, we show how the market maker hedges this position dynamically, and earns some cash due to S_t Oscillations.

4.1. Initial Position and the Hedge

Suppose this market maker buys a call option from a client. [8] The initial position of the market maker is shown in the top portion of Figure 4. It is a standard long call position. The market maker is not an investor or speculator, and this option is bought with the purpose of keeping it on the books and then selling it to another client. Hence, some mechanical procedures should be followed. First, the market maker needs to fund this position. Second, he or she should hedge the associated risks.

We start with the first requirement. Unlike the end investor, market makers never have "money"of their own. The trade needs to be funded. There are at least two ways of doing this. One is to short an appropriate asset in order to generate the needed funds, while the other is to borrow these funds directly from the money market desk. 9 Suppose the second possibility is selected and the market maker borrows C(t) dollars from the money market desk at an interest rate r_t = r. The net position that puts together the option and the borrowed funds is shown in the bottom part of Figure 4.

Now, consider the risks of the position. It is clear from Figure 4 that the long call position funded by a money market loan is risky. If S_t decreases, the position's many times on a given day cannot afford this. The market maker must hedge the risk by taking another position that will offset these possible losses. When S_t declines, a short position in S_t gains. As S_t changes by DS_t , a short position will change by DS_t. Thus , we might think of using this short position as a hedge.

But there is a potential problem. The long call position is described by a curve, whereas the short position in S_t is represented by a line. This means that the responses of C(t) and S_t , to a change In S_t , are not going to be the identical. Everything else being he same, if the underlying changes by DS_t, the change in the option price will be approximately[10]

$$C(t) \cong Cs\ \Delta St \tag{11}$$

The change in the short position on the other hand will wqual _ Δ St . In fact, the net response of the portfolio

$$Vt = \{\text{long } C(t), \text{ short } St\} \tag{12}$$

To a small change in S_t , will be given by the first- order approximation,

$$\begin{aligned} Vt \quad &\cong C_s\, \Delta S_t - S_t \\ &= (C_s - 1)\, S_t < 0 \end{aligned} \tag{13}$$

8. This means that the client has "hit" the bid price quoted by the market maker.
9. The market maker may also wait for some other client to show up aned buy the option back. Market makers have position limits and can operate for short periods without closing open positions.
10. Due to the assumption of everything else being the same, the DS_t and DC(t) should be interpreted within the context of partial differentiation.

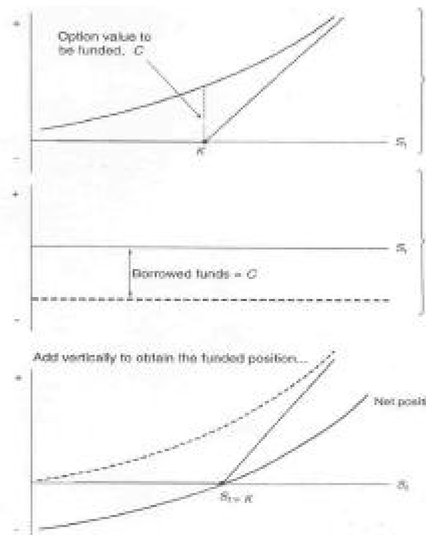


Figure 4

due to the condition 0 < C_s <1. This position is shown in Figure 5. It is still a risky position and, interesting, the risks are reversed. The market maker will now lose money if the S_t increases. In fact, this position amounts to a long put financed by a money market loan.

How can the risks associated with the movements in S_t be eliminated? According to Equation (14), short-selling one unit of S_t overdid the hedge. Instead of short-selling one unit of the S_t asset, the market maker should short h_t units of S_t, selecting the h_t according to

$$h_t = \frac{\partial C(S_t,t)}{\partial S_t} = C_s \tag{15}$$

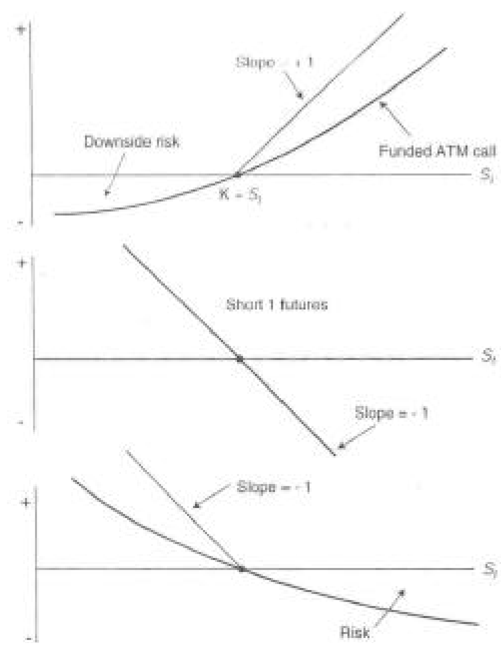


Figure 5

To see why this might work, consider the new portfolio V:

$$V_t = \{\text{long 1 unit of } C(t), \text{ borrow } C(t) \text{ dollars, short } C_s \text{ units of } S_t\} \quad (16)$$

If S_t changes by DS_t, everything else being the same, the change in this porfolio's value will, approximately, be

$$\Delta V_t \cong [C(S_t + \Delta S_t,t) - C(S_t,t)] - Cs\, \Delta S_t \quad (17)$$

We can use a first-order Taylar series approximation of $C(S_t + \Delta S_t, t)$, around point S_t, to simplify this relationship:[11]

[11] Let f(x), be continuous and infinitely differentiable function of x. The kth order Taylor series approximation of f(x), at point x_0, is given by

$f(x) = f(x0)\ d\ f'(x0)\ (x-x0) + f''\ (x0)(x-x0)^2 + \ldots\ldots\ldots\ldots + f^k(x0)\ (x - x0)^k$

where $f^k(x_0)$ is the kth derivative of f(.) evaluated at $x = x_0$

$$C(S_t + \Delta S_t,t) = C(St,t) + \frac{\partial C(St,t)}{\partial St} \Delta St + R \quad (18)$$

Here, R is the remainder. The right-hand side of this formula can be substituted in Equation (17) to obtain

$$\Delta V_t \cong \frac{\partial C(S_t,t)}{\partial S_t} \Delta S_t + R - C_s\, \Delta S_t \quad (19)$$

After using the definition

$$\frac{\partial C(S_t,t)}{\partial S_t} = C_s \quad (20)$$

and simplifying, this becomes

$$\Delta V_t \cong R \quad (21)$$

That is to say, this portfolio's sensitivity towards changes in S_t will be the remainder term, R. It is related to Ito's Lemma, shown in the appendix. The biggest term in the remainder, is given by

$$\frac{1}{2} \frac{\partial^2 C(S_t,t)}{\partial S_t^2} (\Delta St)^2 \quad (22)$$

Since the second partial derivative of C(t) is always positive, the portfolio's value will always be positively affected by small changes in S_t. This is shown in the bottom part of Figure 6. A portfolio such as this one is said to be delta-neutral. That is to say, the delta-exposure, represented by the first-order sensitivity of the position to changes in S_t, is zero. Notice that during this discussion the time variable, t, was treated as constant.

This way of constructing a hedge for options is called delta hedging and the h_t is called the hedge ratio. It is important to realize that the procedure will need constant updating of the hedge ratio, ht, as time passes and, S_t changes. After all, the idea depends on a first-order Taylor series approximation of a nonlinear instrument using a linear instrument. Yet, Taylor series approximations are local and they are satisfactory only for a reasonable neighborhood around the initial St. Consider Figure 8-7. When S_t moves from point A to point B, the approximation at A deteriorates and a new approximation is needed. This new approximation will be the tangent at point B.

4.2 Adjusting the Hedge over Time

We now consider what happens to the delta-hedged position as S_t oscillates. According to our discussion in the previous chapter, as time passes, the replicating portfolio needs to be rebalanced. This rebalancing will generate cash gains.

We discuss these portfolio adjustments in a highly simplified environment. Considering a sequence of simple oscillations in S_t around an initial point $S_{t0} = S^0$, let

$$t_0 < t_1 < \ldots < t_n \quad (23)$$

with

$$t_i - t_{i-1} = \Delta \quad (24)$$

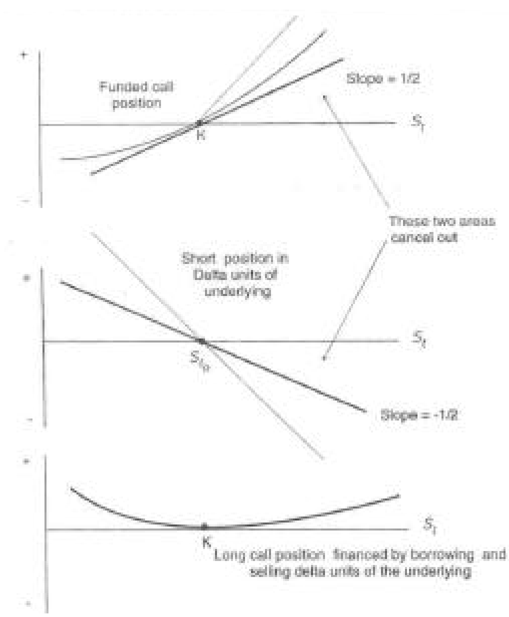


Figure 6

denote successive time periods that are apart Δ units of time. We assume that, S_t oscillates at an annual percentage rate of one standard deviation, s, around the initial point $S_{t0} = S^0$. For example, one possible round turn may be

$$S^0 \rightarrow S^0 + \Delta S \rightarrow S^0 \quad (25)$$

With $\Delta S = sS^0\sqrt{\Delta}$, the percentage oscillations will be proportional to $\sqrt{\Delta}$. The menhanics of maintaining the delta-hedge long call position eill be discssed in this simplified setting.

Since S_{ti} moves between three possible values only, we simplify the notation, and denote the possible values of S_t by S^-, S^0, and S^+, where[12]

$$S^+ = S^0 + \Delta S \quad (27)$$

$$S^- = S^0 - \Delta S \quad (28)$$

12 we can represent this trajectory by a three-state Markov chain that has the following probabilities :

$$P(S0|S+) = 1 \quad P(S-|S0) = \frac{1}{2} \quad P(S+S0) = \frac{1}{2} \quad P(S0|S-) = 1 \quad (26)$$

where S^0 is the initial point.

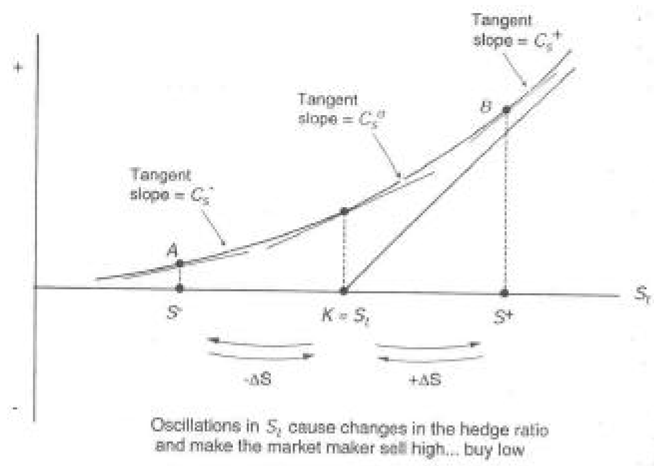


Figure 7

We now show how these oscillations can generate cash gains. According to Figure 8-7, as S_t fluctuates, the slope, C_s, of the $C(S_t, t)$ also changes. Ignoring the effect of time, the slope will change, say, between C_s^+, C^0_s, and C_s^-, as shown in Figure 8-7.[13] We note that

$$C^- < C^0 < C^+ \qquad (29)$$

for all t_i. This means that, as S_t moves, h_t, the hedge ratio will change in a particula way. In order to keep the portfolio delta-hedged, the market maker needs to adjust the number of the underlying S_t that was shorted.

Second, and unexpectedly, the hedge adjustments have a "nice" effect. When S_t moves to S^-, the market maker has to decrease the size of the short position in S_t. To do this, the market maker needs to "buy" back a portion of the underlying asset that was originally shorted at a higher price S^0.

Accordingly, the market maker sells short when prices are high, and covers part of the position and when prices decline. This leads to cash gains.

Consider now what happens when the move is from S^0 to S^+. The new slope, C_s^+, is steeper than the old, C_s^0. This means that the market maker needs to short more of the S_t-asset at the new price. What the S_t moves back to S^0, these shorts are covered at S^0, which is lower than S^+. This again leads to cash gains.

[13] It is important to realize that these slope also depend on time t, although, to simplfy the notation, we are omitting the time index here.

Thus, as S_t oscillates around S_0, the portfolio is adjusted accordingly, and the market maker would, automatically, sell highly and buy low. At every round turn, say, $\{S^0, S^+, S^0\}$, which takes two periods, the hedge adjustments will generate a cash gain equal to

$$(C^+ - C^0)\,[\,S^0 + S) - S^0] = (C^+ - C^0)\,\Delta S_t \qquad (30)$$

Here, the $(C^+ - C^0)$ represents the number of S_t asset that were shorted after the price moved from S^0 to S^+. Once the price goes back to S^0, the same securities are purchased at a lower price. It is interesting to look at these trading gains as the time interval, Δ, becomes smaller and smaller.

4.2.1. Limiting From

As $\Delta S \to 0$, we can show an important approximation to the trading (hedging) gains

$$(C^+ - C^0)\,\Delta S \qquad (31)$$

The term $(C^+ - C^+)$ is the change in the first partial derivative of $C(S_t, t)$, as S_t moves from S_{t0} to a new level denoted by $S_{t0} + \Delta S$. We can convert the $(C^+ - C^0)$ into a rate of change after multiplying and dividing by ΔS:

$$(C^+ - C^0)\,\Delta S = \frac{C^+ - C^0}{\Delta S}(\Delta S)^2 \qquad (32)$$

As we let ΔS go to zero, we obtain the approximation

$$\frac{C^+ - C^0}{\Delta S} \to \frac{\partial^2 C(St,t)}{\partial S^2 t} \qquad (33)$$

Thus, the round-turn gains from delta-hedge adjustments shown in Equation (30) can be approximated as

$$(C^+s - C^0s) \cong \frac{\partial^2 C(St,t)}{\partial S^2 t}(\Delta S)^2 \qquad (34)$$

Per time unit gains are then half of this,

$$\frac{1}{2}\frac{\partial^2 C(St,t)}{S^2 t}(\Delta S)^2 \qquad (35)$$

These gains are only part of the potential cash inflows and outflows faced by the market maker. The position has further potential cash flows that need to be described. This is done on the next two sections.

4.3. Other Cash Flows

We just showed that oscillations in St generate positive cash flows if the market maker delta hedges his or her long option position. Does this imply an arbitrage opportunity? After all, the market maker did not advance any cash, yet seems to receive cash spontaneously, as long as St oscillates. The answer is no. There are costs to this strategy, and the delta – hedged option position is not riskless.

1. The market maker funded his or her position with borrowed money. This means, that , as time passes, an interest cost is incurred. For a period of length Δ, this cost will equal

$$rC\Delta \qquad (36)$$

Under the constant spot rate assumption. (We write C(t), as C.)

2. The option has time value, and, as time passes, everything else being the same the value of the option will decline at the rate

$$Ct = \frac{\partial C(St,t)}{\partial t} \qquad (37)$$

The option will lose

$$Ct = \frac{\partial C(St,t)}{\partial t} \Delta \quad (38)$$

Dollars, for each Δ that passes.

3. Finally, the cash received from the short position generates $rS_tC_s \Delta$ dollars interest, every time period Δ.

The trading gains and the short position generates generates $rS_t C_8 \Delta$ dollars interest, every time period Δ.

The trading gains and the costs can be put together to obtain an important partial differential equation (PDE), which plays a central role in financial engineering.

4.4 Option Gains and Losses as a PDE

We now add all gains and costs perunit of time Δ. The options' gains per time unit from hedging adjustment is

$$\frac{1}{2}\frac{\partial^2 C(St,t)}{\partial S^2 t}(\Delta S)^2 \quad (39)$$

In case the process St is geometric, the annual percentage variance will be constant and this can be written as (see the Appendix)

$$\frac{1}{2} C_{ss}\sigma^2 S^2_t \Delta \quad (40)$$

The rest of the argument will continue with the assumption of a constant –

Interest is paid daily on the funds borrowed to purchase the call. For every period of length Δ, a long call holder will pay

$$rC\Delta \quad (41)$$

Another item is the interest earned from cash generated by shorting Cs units of St:14

$$\frac{1}{2} C_{ss}\sigma^2 S^2_t \Delta \quad (42)$$

14 If the underlying asset is not "cash" but a futures contract, then this item may drop.

Adding these, we obtain the net cash gains (losses) from the hedged long call position during Δ:

$$\frac{1}{2} C_{ss}\sigma^2 S^2_t \Delta + rC_sS_t \Delta - rC\Delta \quad (43)$$

Now, in order for there to be no arbitrage oppoutunity, this must be equal to the daily loss of time value:

$$\frac{1}{2} C_{ss}\sigma^2 S^2_t \Delta + rC_sS_t \Delta - rC\Delta = -C_t \Delta \quad (44)$$

We can eliminate the common ^ terms, and obtain a very important relationship that some readers will recognize as the Black-Scholes partial differential equation:

$$\frac{1}{2} C_{ss}\sigma^2 S^2_t \Delta + rC_sS_t - rC + Ct = 0 \quad (45)$$

Every PDE comes with some boundary conditions and this is no exception. The call option will expire at time T , and the expiration C(St, T) is given by

$$C(S_T,T) = \max [S_T - K,0]$$

Solving this PDE gives the Black-Scholes equation. In most finance texts, the PDE derived here is obtained from some mathematical derivation. In this section, we obtained the same PDE heuristically , from practical trading and arbitrage arguments.

4.5 Cash Flows at Expiration

The cash flows at expiration date have three components: (1) the market maker has to pay the original loan if it is not paid off slowly over the life of the option, (2) there is the final option settlement, and (3) there is the final payoff from the short S_t position.

Now , at an infinitesimally short time period, dt, before expiration, the price of the underlying will be very close to S_t. Call it S^-T. The price curve $C(S_t, t)$ will be very near the piecewise linear option payoff. Thus, the hedge ratio h_T = Cs will be very close to either zero, or one:

$$h^-_T \begin{cases} 1 & S^-_T > K \\ 0 & S^-_T < K \end{cases} \quad (46)$$

This means that, at time T, any potential gains from the long call option position will be equal to losses on the short S_t position.

The interesting question is, how does the market maker manage to payback the original loan under these conditions? There is only one way. The only cash that is available is the accumulation of (net) trading gains from hedge adjustments during [t, T]. As long as Equation (45) is satisfied for every t_i, the hedged long option position will generate just enough cash to pay back the loan. The option price, C(t), regarded this way is the discounted sum of all gains and losses from a delta- hedged option position the trader will incur based on expected S_t – volatility.

We will now consider a numerical example to our highly simplified discussion of how realized volatility is converted into cash via an option position.

4.5 An Example

Consider a stock, S_t, trading at a price of 100. The stock pays no dividends and is known to have a Black-Scholes volatility of s = 45% per annum. The risk- free interest rate is 4% and the St is known to follow a geometric process, so that the Black – Scholes assumptions are satisfied.

A market maker buys 100 plain vanila, at –the – money calls that expire in 5 days. The premium for one call is 2.13 dollars. This is the price found by plugging the above data into the Black-Scholes formula. Hence, the total cash outlay is $213, buys the call options, and immediately hedges the long position by short selling an appropriate number of the underlying stock.

Example

Suppose that during these 5 days the underlying stock follows the path:

{Day 1 = 100, Day2 = 105, Day3 = 105, Day 5 = 100} (47)

What are the cash flows, gains, and losses generated by this call option that remain on the market maker's books? We answer this below.

1- Day 1 : The purchase date

Current Delta : 51 (found by differentiating the Black-Scholes formula with respect to S_t, plugging in the data and then multiplying by 100.)

Cash paid for the call options :- $213

Amount borrowed to pay for the calls : $213

Amount generated by short selling 51 units of the stock : $5100. This amount is deposited at a rate of 4%.

2- Day 2 :- Price goes to 105

Current Delta : 89 (Evaluated at St= 105, 3 days to expiration)

Interest on amount borrowed: 213(.04) (1/360) = $.02

Interest earned from deposit: 5100(.04) (1/360) = $.57 (Assuming no bid-ask difference in interest rates.)

Short selling 38 units of additional stock to reach delta-neutrality which generate :- 38(105) = $ 3990.

3- Day 3 :- Price goes back to 100

Current Delta : 51

Interest on amount borrowed : 213 (.04) ( ) = $.02

Interest earned from deposits : (5100 + 3900) (.04) = $1.

Short covering 38 units of additional stock at 100 each, to reach delta neutrality generates a cash flow of : 38(5) = $190. Interest on these profits is ignored to the first order of approximation.

4- Day 4 : Price goes to 105

Current Delta : 98

Interest on amount borrowed : 213(.04) (1/360) = $.02

Interest earned from deposits : 5100(1/360) = $.57

Shorting 47 units of additional stock at 105 each, to reach delta neutrality generates: 47 (105)= $4935.

5- Day 5 : Expiration with St= 100

Net cash generated from covering the short position: 47(5) = $235 (There wee 98 shorts, covered at $100 each. 47 shorts were sold at $ 105. 51 Shorts at $100).

Interest on amount borrowed : 213(.04) (1/360) = $.02

Interest earned from deposits (5100 + 4935)(.04) (1/360) = $1.1. The option expires at-the-money and generates no extra cash.

6- Totals

Total interest paid : 4(02) = $.08

Total interest earned: 2(.57) + 1 + 1.1 = $3.24

Total cash earned from hedging adjustments: $235 + $190

Cash needed to repay the loan: $213

Total net profit ignoring interest on interest: $213.16.

A more exact calculation would take into account interest on interest earned and the interest earned on the $190 for 2 days.

We can explain why total profit is positive. The path followed by St in this example amounts to a daily actual volatility of 5%. Yet, the option was sold at an annual implied volatility of 45% which corresponds to a "daily" percentage implied volatility of:

$$0.45 \sqrt{1/360} = 2.36\% \qquad (48)$$

Hence, during the life of options, the St fluctuated more than what the implied volatility suggested. As a result, the long convexity position had a net profit.

This example is, of course, highly simplified. It keeps implied volatility constant and the oscillations occur around a fixed point. If these assumptions are relaxed, the calculations will change.

4.5.1 Some Caveats

Three assumptions simplified notation and discussion in this section.

- First, we considered oscillations around a fixed S^0. In real life, oscillations will clearly occur around points that themselves move. As this happens, the partial derivatives, C^s and C^{ss}, will change more complicated ways.
- Second, C_s and C_{ss} are also functions of time t, and as time passes, this will be another source of change.
- The third point is more important. During the discussion, oscillations were kept constant at Δ S. In real life, volatility may change over time and be random as well. This would not invalidate the essence of our argument concerning gains from hedge adjustments, but it will clearly introduce another risk that the marker may have to hedge against. This risk is known as mega risk.
- Finally, it should be remembered that the underlying asset did not make any payouts during the life of the option. If dividends or coupons are paid, the calculation of cash gains and losses needs to be adjusted accordingly.

These assumptions were made to emphasize the role of options as volatility instruments, forthcoming chapters will deal with how to relax them.

5. Tools for Options

The Black–Scholes PDE can be exploited to obtain the major tools available to an another trader or market maker. First of these is the Black-Scholes formula, which gives the arbitrage-free price of a plain vanilla call (put) option under specific assumptions.

The second set of tools is made up of the "Greeks". These measure the sensitivity of an option's price with respect to changes in various parameters. The Greeks are essential in hedging and risk managing options books. They are also used in pricing and in options strategies.

The third set of tools are ad-hoc modifications of these theoretical constructs by market practitioners. These modifications adapt the theoretical tools to the real world, making them more "realistic".

5.1 Solving the Fundamental PDE

The convexity of options payoffs implies argument, namely that the expected net gains(losses) from St oscillations are equal to time decay during the same period. This leads to the Black-Scholes PDE.

$$\frac{1}{2} C_{ss}\sigma^2 S_t^2 + rC_s S_t - rC + C_t = 0 \qquad (49)$$

with the boundary condition

$C(T) = \max [S_T - K, 0]$ (50)

Now, under some conditions partial differential equations can be solved analytically and a closed-form formula can be obtained. See Duffie (2001). In our case, with specific assumptions concerning the dynamics of S_t, this PDE has such a closed-form solution. This solution is the market benchmark known as the Black-Scholes formula.

1. The risk-free interest rate is constant at r.
2. The underlying stock price dynamics are described in continuous time by the stochastic differential equation (SDE)1:[15]

$dS_t = \mu (S_t) S_t dt + \sigma S_t dW_t \; [0,\infty]$ (51)

where Wt represents a Wiener process with respect to real-world probability P.[16]

15 The appendix to this chapter discusses SDEs further,

16 The assumptions of Wiener process implies heuristically that

$Et[dWt] = 0$ (52)

and that

$Et[dWt]^2 = dt$ (53)

These increments are the continuous time equivalent of sequences of normally distributed variables. For discussion of stochastic differential equations and the Wiener process, see, for example, Oksendal (2003), Neftci(2000) provides the heuristics.

To emphasize an important aspect of the previous SDE, the dynamics of St, are assumed to have a constant percentage variance during infinitesimally short intervals. Yet, the drift component, $\mu (S_t)S_t$, can be general and need nor be specified further. Arbitrage arguments are used to eliminate the $\mu (S_t)$ and replace it with the risk-free instantaneous spot rate r in the previous equation.

3. The stock pays no dividends, and there are no stock splits or other corporate actions during the period [t,T].
4. Finally, there are no transaction costs and no bid-ask spreads.

Under these assumptions, we can solve the PDE in equations (49) and (51) and obtain the Black-Scholes formula:

$C(t) = S_t N(D_1) - Ke^{-r(T-t)} N)(d_2)$ (54)

where d1,d2 are

$$d_1 = \frac{\log (St/k) + (r + \sigma^2/2) (T-t)}{\sigma\sqrt{T-t}} \quad (55)$$

$$d2 = \frac{\log(St,k) + (r - \sigma 2/2) (T-t)}{\sigma\sqrt{T-t}} \quad (56)$$

The N(x) denotes the cumulative standard normal probability:

$$N(x) = \frac{1}{2\pi} e - ½ u2 du \quad (57)$$

In this formual r, s ,T, and K are considered parameters, since the formula holds in this version only when these components are kept constant. 17 The variable are St and t. The latter is allowed to change during the life of the option.

Given this formula, we can take the partial derivatives of,

$C(t) = C(S_t, t | r, \sigma ,T, K)$ (58)

with respect to the variables St and t and with respect to the parameters r, s ,T, and K. These partials are the Greeks. They represent the sensitivities of the option price with respect to a small variation in the parameters and variables.

We will study the Blach Scholes Formula in the next lesson as a continuation of this lesson. You read it so that we can discuss it nicely.

BLACK SCHOLES MODEL

Objectives

- After completion of this lesson you will be able to identify fully of how the options market players work and at the same time to what extent you are exposed to such volatility and of how BS Model can be helpful in such conditions.

This lesson is the continuation of lesson number 29. all the symbols we use here and its sequencing is done as per the continuity of the previous one.

Black's Formula/Model

The Black-Scholes formula in Equation(54) is the solution to the fundamental PDE when delta hedging is done with the "cash" underlying. As discussed earlier, trading gains and funding costs lead to the PDE.

$$rC_sS_t - rC + \frac{1}{2} C_{ss}\, 2S2t = -Ct \qquad (59)$$

with the boundary condition:

$$C(S_T,T) = \max[S_T - K,0] \qquad (60)$$

17 The Volatility needs to be constant during the life of the option. Otherwise, the formula will not hold, even though the logic behind the derivation would.

When the underlying becomes a forward contract, the S_t will become the corresponding forward price denoted by F_t and the Black-Scholes PDE will change slightly.

Unlike a cash underlying, buying and selling a forward contract does not involve funding. Long and short forward positions are commitments to buy and sell at a future date T, rather than outright purchases of the underlying asset. Thus, the only cash movements will be interest expense for funding the call, and cash gains from hedge adjustments. The means that the corresponding PDE will look like

$$-rC + \frac{1}{2} C_{ss}\, \sigma^2 F^2_t = -C_t \qquad (61)$$

with same boundary condition :

$$C(F_T,T) = \max [F_T - K, 0] \qquad (62)$$

where F_t is now the forward price of the underlying.

The solution to this PDE is given by the so-called Black's formula in the case where, the options are of European style.

$$C(F_t,t)^{Black} = e^{-r(T-t)} [F_tN(d_1) - KN(d_2)] \qquad (63)$$

with

$$d_1^{Black} = \frac{\log Ft/K + ½\, \sigma^2 (T-t)}{\sigma \sqrt{(T-t)}} \qquad (64)$$

$$d_2^{Black} = d_1^{Black} - \sigma \sqrt{(T-t)} \qquad (65)$$

Black's formula is useful in many practical circumstances where the Black-Scholes formula cannot be applied directly. Interest rate derivatives such as caps and floors for example, are options written on Libor rates that will be observed at future dates. Such settings lend themselves better to the use of Black's formula. The underlying risk is a forward interest rate such as forward Libor, and the relate option prices are given by Black's formula. However, the reader should remember that in the preceding version of Black's formula the spot rate is taken as constant. In chapter 15 this assumption will be relaxed.

5.3 Other Formulas

The Black-Scholes type PDEs can be solved for a closed-form formula under somewhat different conditions as well. These operations result in expressions that are similar but contain further parameters and variables. We consider two cases of interest. Our first example is a chooser option.

5.3.1 Chooser Options

Consider a vanilla put, P(t) and a vanilla call, C(t) written on S_t with strike K, expiration T. A chooser option then is an option that gives the right to choose between C(t) and P(t) at some later date T_0. Its payoff at time T_0, with T_0<T is

$$C^h (T_0) = \max [C(S_{T0}, T_0), P(S_{T0},T_0)] \qquad (66)$$

Arbitrage arguments lead to the equality

$$P(S_{T0},T0) = - (S_{T0} - Ke\text{-}r(T\text{-}T0) + C(S_{T0},T0) \qquad (67)$$

Using this, (66) can be written as

$$C^h (T_0) = \max[C(S_{T0},T_0), - (S_{T0} - Ke\text{-}r(T\text{-}T0) + C(S_{T0}, T_0) \qquad (68)$$

or taking the common term out,

$$C^h (T_0) = C(S_{T0} ,T_0), + \max[-S_{T0} - K^{e-r(T-T0)} ,0] \qquad (69)$$

In other words, the chooser option payoff is either equal to the value of the call at time T_0, or it is that plus a positive increment, in the case that

$$(S_{T0} - Ke^{-r(T-T0)}) < 0 \qquad (70)$$

But, this is equal to the payoff of a put with strike price $Ke^{-r(T-T0)}$ and exercise date T_0. Thus, the pricing formula for the chooser option is given by

$$C^h(t) = [S_tN(d_1) - Ke^{-r(T-t)} N (d_2) +[-S_t N (-d_1) + Ke^{-r(T-T0)} e^{-r(T0-T)} N (-d_2)] \qquad (71)$$

Simplifying

5.4. Uses of Block Scholes Type Formulas

Obviously, the assumptions underlying the derivation of the Black-Scholes formula are quite restrictive. This becomes especially clear from the way we introduced options in this book. In particular, if options are used to bet on the direction of volatility, then how can the assumption of constant percentage volatility possibly be satisfied? This issue will be discussed further in later chapters where the way market professionals use the Black-Scholes formula while trading, volatility is clarified.

When the underlying asset is an interest rate instrument or a foreign currency, some of the Black-Scholes assumptions become untenable. [18] Yet, when these assumptions are relaxed, the logic used in deriving the Black-Scholes formula may not result in a POE that can be solved for a closed-form formula.

Hence, a market practitioner may want to use the Black-Scholes formula or variants of it, and then adjust the formula in some ad-hoc yet, practical ways. This may be preferable to trying to derive new complicated formulas that may accommodate more realistic assumptions. Also, even though the Black-Scholes formula does not hold when the underlying assumptions change, acting as if the assumptions hold yields results that are surprisingly robust [l9] We will' see that this is exactly what happens when traders adjust the volatility parameter depending on the "moneyless" of the option under consideration.

This completes our brief discussion of the first set of tools that are essential for option analysis, namely Black-Scholes types closed-form formulas that give the arbitrage-free price of an option under some stringent conditions. Next, we discuss the second set of toots that traders and market makers routinely use. These are various sensitivity factors called the Greeks.

6. The Greeks and Their Uses

The Black-Scholes formula gives the value of a vanilla call (put) option under some specific assumptions. Obviously, this is useful for calculating the arbitrage-free value of an option. But, a financial engineer needs methods for determining how the option premium, C(t), changes as the variables or the parameters in the formula change within the market environment This is important since the assumptions used in deriving the Black-Scholes formula are unrealistic. Traders, market makers, or risk managers must constantly monitor the sensitivity of their option books with respect to changes in S_t, r, t, or s .

Example

The case of a change in s is a good example. We motivated option positions essen-tially (but not fully) as positions taken on the volatility. It is clear that volatility is not constant as assumed in the Black-Scholes world. Once an option is bought and delta hedged, the hedge ratio C_S and the C_{SS} both depend on the movements in the volatility parameters.

Hence, the "hedged" option position will still be risky in many ways. For example depending on the way changes in s and S_t affect the C_{SS}, a market maker may be correct in his or her forecast of much S_t will fluctuate, yet may still lose money on a long option position.

A further difficulty is that option sensitivities may not he uniform across the strike price K or expiration T. For options written on the same underlying, differences in K and T lead to what are called smile effects and term structure effects, respectively, and should be taken into account carefully.

Option sensitivity parameters are called the "Greeks" in the options literature. We discuss them next and provide several practical examples.

6.1. Delta

Consider the Black-Scholes formula $C(S_t, t\ r, s\ , T, K)$. How much would this theoretical price change if the underlying asset price, S_t, moved by an infinitesimal amount?

One theoretical answer to this question can be given by using the partial derivative of the function with respect to S_t. This is by definition the delta at time t:

$$\text{Delta} = \frac{\partial C\langle S_t t \mid r, s\ , T, K\rangle}{\partial S_t} \tag{81}$$

This partial derivative was denoted by C_S earlier. Note that delta is the local sensitivity of the option price to an infinitesimal change in S_t only, which incidentally is the reason behind using partial derivative notation.

To get some intuition on this, remember that the price curve for a long call has an upward slope in the standard C(t), S_t space. Being the slope of the tangent to this curve, the delta of .a long call (put) is always positive (negative). The situation is represented in Figure 8-9. Here, we consider three outcomes for the underlying asset price represented by S_A, S_B, and S_C and hence obtain three points, A, B, and C. on the option pricing curve. At each point. we can draw a tangent. The slope of this tangent corresponds to the delta at the respective price.

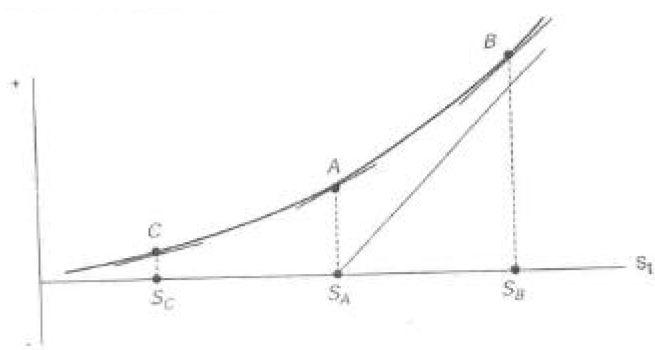


Figure 8

- At point C, the slope, and hence, the delta is close to zero, since the curve is approaching the horizontal axis as S_t falls.
- At point B, the delta is close to one, since the curve is approaching a line with slope + 1.-
- At point A, the delta is in the "middle," and the slope of the tangent is between zero and one.

Thus, we always have 0 < delta < 1 in case of a long call position. As mentioned earlier, when the option is at-the-money (ATM), the delta is close to. 5.

6.1.1. Convention

Market professionals do not like to use decimal points. The convention in option markets is to think about trading not one, but 100 options, so that the delta of option positions can be referred to in whole numbers, between 0 and 100. According to this convention, the delta of an ATM option is around 50.A25-delta option would be out-of-the-money and a 75-delta option in-the-money. Especially in FX markets, traders use this terminology to trade options.

Under these conditions, an options trader may evaluate his or her exposure using delta points. A trader may be long delta, which means that the position gains if the underlying increases, and loses if the underlying decreases. A short delta position implies the opposite.

6.1.2. The Exact Expression

The partial derivative in Equation (81) can be taken in case the call option is European and the price is given by the Black-Scholes formula. Doing so, we obtain the delta of this important special case:

$$\frac{\partial C(S_t,t,r,\sigma,T,K)}{\partial S_t} = \int_{-\infty}^{\frac{(T-t)(r+\frac{1}{2}\sigma^2)+\log(S_t/K)}{\sqrt{(T-t)}\,\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\,dx \quad (82)$$

$$= N(di)$$

This derivation is summarized in the appendix to this chapter. It is shown that the delta is itself a function that depends on the "variables" S_t, K, r, s , and on the remaining life of the option, T –t. This function is in the form of a probability. The delta is between 0 and I, and the function will have the familiar S-shape of a continuous cumulative distribution function (CDF). This, incidentally, means that the derivative of the delta with respect to S_t, which is called gamma, will have the shape of a probability density function (PDF).[20] A typical delta will thus look like the S-shaped curve shown in Figure 10.

We can also see from this formula how various movements in market variables will affect this particular option sensitivity. The formula shows that whatever increases the ratio

$$\frac{\log(S_t/K)+(r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad (83)$$

will increase the delta; whatever decreases this ratio, will decrease the delta.

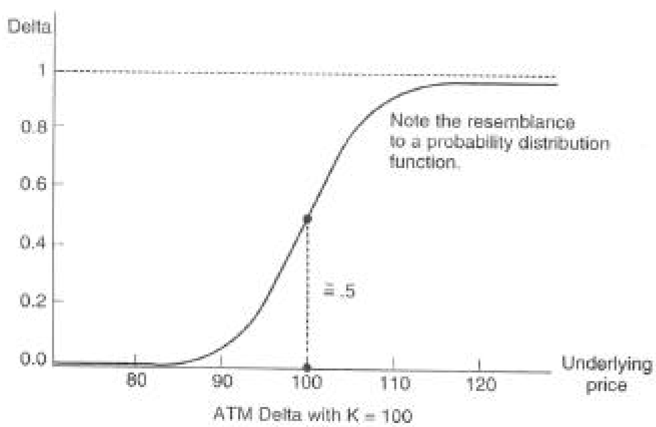


Figure 9

For example, it is clear that as r increases, the delta will increase. On the other hand, a decrease in the moneyness of the call option, defined as the ratio

$$\frac{S_t}{K} \quad (84)$$

decreases the delta. The effect of volatility changes is more ambiguous and depends on the moneyness of the option.

Example

We calculate the delta for some specific options. We first assume the Black-Scholes world, even though the relevant market we are operating in may violate many of the Black-Scholes assumptions. This assumes, for example, that the dividend yield of the underlying is zero and this assumption may not be satisfied in real life cases. Second, we differentiate the function C(t)

$$C(t) = S_t N(d_1) - K_e^{-r\,(T-t)}\, N(d_2) \quad (85)$$

where the d1 and d2 are as given in Equations (55)-(56), with respect to S_t. Then, we substitute values observed for S_t, K, r, s , (T - t). -

Suppose the Microsoft December calls and puts shown in the table from our first example, in this chapter satisfies these assumptions. The deltas can be calculated based on the following parameter values:

$$St = 61.15,\ r = .025,\ \sigma = 30.7\%,T - t = 58/365 \quad (86)$$

Here, σ is the implied volatility obtained by solving the equation for K = 60

$$C(61.15.60..025.58/365,\sigma) = \text{Observed price} \quad (87)$$

Plugging the observed data into the formula for delta yields the following values:

Calls	Delta
Dec 55.00	.82
Dec 60.00	.59
Dec 65.00	.34
Dec 70.00	.16

Puts	Delta
Dec 55.00	-.17
Dec 60.00	-.40
Dec 65.00	-.65
Dec 70.00	-.84

We can make some interesting observations:

1. The ATM calls and puts have the same price.
2. Their deltas, however, are different.
3. The calls and puts that are equally far from the ATM, have slightly different deltas in absolute value.

According to the last point, if we consider 25-delta calls and puts, they will not have exactly the same.[21]

We now point out to some questionable assumptions used in our example. First, In calculating the deltas for various strikes, we always used the same volatility parameter s. This is a not a trivial point. Options that are identical in every other aspect, except for their strike K may have different implied volatilities. There may be a volatility smile. Using the ATM implied volatility in calculating the delta of all options may not be the correct procedure. Second, we assumed a zero dividend yield, which is not realistic either. Normally, stocks have positive expected dividend yields and some correction for this should be made when option prices and the relevant Greeks are calculated.

A rough way of doing it, is to calculate an annual expected percentage dividend yield and subtract it from the risk-free rate r.

6.2. Gamma

Gamma represents the rate of change of the delta as the underlying risk S_t changes. Changes in delta were seen to playa fundamental role in determining the price of a vanilla option. Hence, gamma is another important Greek. It is given by the second partial derivative of $C(S_{t,}\ t\ )$with: respect to S_t:

$$\text{Gamma} = \frac{\partial^2 C(S_t, t | r, \sigma, T, K)}{\partial S_t^2} \qquad (88)$$

We can easily obtain the exact expression for gamma in the case of a European call. The derivation in the appendix gives

$$\frac{S^2C(S_t, tr, s, T, K)}{dS_t^2} = S_t s\sqrt{T-t} \times \left[\log(S/K)^{-1/2} + r[T-t]^{1/2} + s^2(T - \sqrt{2}p/e)\right] \qquad (89)$$

Gamma shows how much the delta hedge should be adjusted as S_t changes. Figure 8-11 illustrates the gamma for the Black-Scholes formula. We see the already-mentioned property. Gamma is highest, if the option is at-the-money, and approaches zero, as the option becomes deep in-the- -money or out-of-the-money.

We can gain some intuition on the shape of the gamma curve. First, remember that gamma is, in fact, the derivative of the delta with respect to S_t. Second, remember that delta itself had the shape of a cumulative normal distribution. This means that the shape of gamma will be similar to that of a continuous, bell-shaped probability density function, as expression (89) indicates.

Consider now a numerical example dealing with gamma calculations. We use the same data utilized earlier in the chapter.

Example

To calculate the gamma, we use the same table as ill the firs example in the chapter. We take the partial derivative of the delta with respect to S_t. This gives a new function of

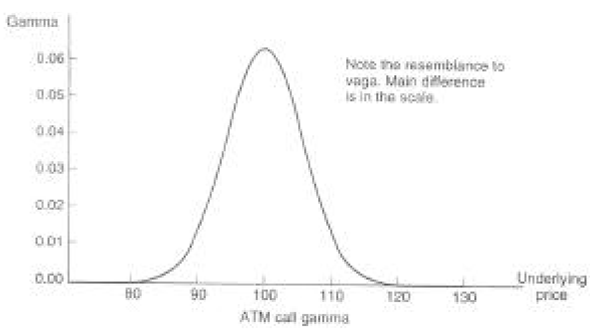


Figure 10

S_t, K, T, s , (T - t), which measures the sensitivity of the delta to the underlying S_t. We then substitute the observed values for the S_t, K, r, s , (T - t) to obtain the gamma at that particular point.

For the Microsoft December calls and puts shown in the table, the gammas are calculated based on the parameter values.

$$S_t = 60.0,\ r = .025,\ \sigma = .31\%,\ T - t = 58/365,\ k = 60 \qquad (90)$$

where s if the implicit volatility.

Again we are using the implicit volatility that corresponds to the ATM option in calculating the delta of all options, in-the-money or out.

Plugging the observed data into the formula for gamma yields the following values:

Calls	Gamma
Dec 55.00	.034
Dec 60.00	.053
Dec 65.00	.050
Dec 70.00	.032
Puts	**Gamma**
Dec 55.00	.034
Dec 60.00	.053
Dec 65.00	.050
Dec 70.00	.032

The following observations can be made:

1. The puts and calls with different different to the ATM strike have gammas that are alike but not exactly symmetric.
2. Gamma is positive if the market maker is long the option; otherwise it is negative.

It is also clear from this table that the gamma is highest when we are dealing with an ATM option.

Finally, we should mention that as time passes, the second-order curvature of ATM options will increase as the gamma function becomes more peaked and its tails will go toward zero.

6.2.1. Market Use

We must comment on the role played by gamma in option trading. We have seen that long delta exposures can be hedged by going short using the underlying asset. But, how are gamma exposure hedged? Traders sometimes find this quite difficult. Especially in very short-dated, deep out-of-the-money options, gamma can suddenly go from zero to very high values and may cause significant losses (or gains).

Example

The forex option market was caught short gamma in GBP/ EUR last week. The spot rate surged from GBPO.6742 to GBPO.6973 late the previous week; one-month volatilities went up from about 9.6% to roughly 13.3%. This move forced players to cover their gamma. (A typical market quote.)

This example shows one way delta and gamma are used by market professionals. Especially in the foreign exchange markets, options of varying moneyless characteristics are labeled accord-ing to their delta. For example, consider 25-delta Sterling puts. Given that an at-the-money put has a delta of around 50, these puts are out-of-the-money. Market makers had sold such options and, after hedging their delta exposure, were holding short gamma positions. This meant that as the Sterling-Euro exchange rate fluctuated, hedge adjustments led to higher than expected cash outflows.

6.3. Vega

A critical Greek is the vega. How much will the value of an option change, if the volatility parameter, s , moves by an infinitesimal amount? This question relates to an option's sensitivity with respect to implied volatility movements. Vega is obtained by taking the partial derivative of the function with respect to s :

$$\text{Vega} = \frac{\partial C(St, t, r, s, T, K)}{\partial s} \tag{91}$$

An example of vega is shown in Figure 8-12 for a call option. Note the resemblance to the gamma displayed earlier in Figure 8-11. According to this figure, the vega is greatest when the option is at-the-money. This implies that, if we use the ATM option as a vehicle to benefit from oscillations in S_t. we will also have maximum exposure to movements in the implied volatility. We consider some examples of vega calculations using actual data.

Example

Vega is the sensitivity with respect to the percentage volatility parameter, s , of the option. According to, the, convention, this is calculated using the Black-Scholes formula. We differentiate the formula with respect to the volatility parameters.

Doing this and then substituting

$$C(61.015, .025, 60, 58/365, \sigma) = \text{Observed price} \tag{92}$$

We get a measure of how this option's prices will react to small changes in s .

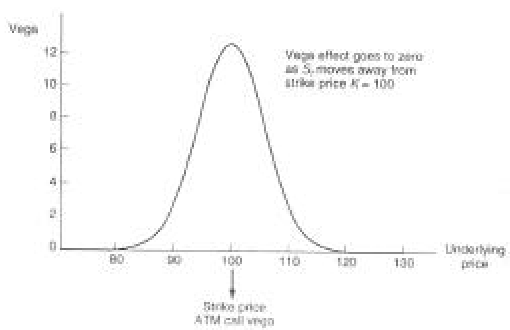


Figure 11

For the table above, we get the following results:

Calls	Vega ($)
Dec 55.00	6.02
Dec 60.00	9.4
Dec 65.00	8.9
Dec 70.00	5.6
Puts	**Vega ($)**
Dec 55.00	6.02
Dec 60.00	9.4
Dec 65.00	8.9
Dec 70.00	5.6

We can make the following comments:

1. At-the-money options have the largest values of vega.
2. As implied volatility increases, the ATM vega changes marginally, whereas the out-of-the-money option vegas do change, and in the same direction.

Option traders can use the vege in calculating the "new" option price in case implied volatilities change by some projected amount. For example, in the preceding example, if the implied volatility increases by 2 percentage points, then the value of the Dec 60-put will increase approx-imately by 0.19, everything else being the same.

6.3.1. Market Use

Vega is an important Greek because it permits market profes-sionals to keep track of their exposure to changes in implied volatility. This is important, since the Black-Scholes formula is derived in a framework where volatility is assumed to be constant, yet, used in an environment where the volatility parameter, s , changes. Market makers often quote the s directly, instead of quoting the Black-Scholes value of the option. Under these conditions, vega can be used to track exposure of option books to changes in the s . This can be followed by vega hedging.

The following reading is one example of the use of vega by the traders.

Example

Players dumped USD/JPY vol last week in a quite spot market, causing volatilities to go down further. One player was selling USD l billion in six-month dollar/yen options in the market. These trades were entered to hedge vega exposure. The drop in the vols forced market makers to hedge exotic trades they had previously sold.

According to this reading, some practitioners were long volatility. They had bought options when the dollar-yen exchange rate volatility was higher. They faced vega risk. If implied volatility declined, their position would lose value at a rate depending on the position's vega. To cover these risky positions, they sold volatility and caused further declines in this latter. The size of vega is useful in determining such risks faced by such long or short volatility posi-tions.

6.3.2. Vega Hedging

Vega is the response of the option value to a change in implied volatility. In a liquid market, option traders quote implied volatility and this latter continuously fluctuates. This means that the value of an existing option position also changes as implied volatility changes. Traders who would like to eliminate this exposure use vega hedging in making their portfolio vega-neutral. Vega hedging in practice involves buying and selling options, since only these instruments have convexity and hence, have vega.

6.4. Theta

Next, we ask how much the theoretical price of an option would change, if a small amount of, time, dt, passes. We use the partial derivative of the function with respect to time parameter t, which is called theta:

$$\text{Theta} = \frac{\partial C(S_t, t, r, s, T, K)}{\partial t} \tag{93}$$

According to this, theta measures the decay in the time value of the option. The intuition behind theta is simple. As time passes, one has less time to gain from future S_t oscillations. Option's time value decreases. Thus, we must have theta < O.

If the Black-Scholes assumptions are correct, we can calculate this derivative analytically and plot it. The derivative is represented in Figure 8-13. We see that, all else being the same a plain vanilla option's time value will decrease at a faster rate as expiration approaches.

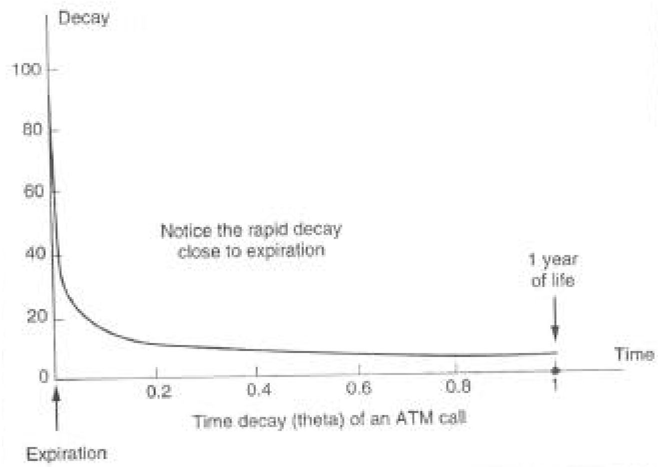


Figure 12

6.5. Omega

This Greek relates to American options only and is an approximate measure developed by market professionals to measure the expected life of an American-style option.

6.6. Higher-Order Derivatives

The Greeks seen thus far are not the only sensitivities of interest. One can imagine many other sensitivities that are important to market professionals and investors. In fact, we can calculate the sensitivity of the previously mentioned Greeks themselves with respect to S_t, s , t, and r. These are higher-order cross partial derivatives and under some circumstances will be quite relevant to the trader.

Two examples are as follows. Consider the gamma of an option. This Greek determines how much cash can be earned as the underlying S_t oscillates. But the value of the gamma depends on the S_t and s as well. Thus, a gamma trader may be quite interested in the following sensitivities:

$$\frac{\partial \text{Gamma}}{\partial S_t} \frac{\partial \text{Gamma}}{\partial s} \qquad (94)$$

These two Greeks are sometimes referred to as the speed and volga, respectively, It is obvious that the magnitude of these partials will be useful in determining the risks and gains of gamma positions. Exotic option deltas and gammas may have discontinuities, and such high-order moments may be very relevant.

Another interesting Greek is the derivative of vega with respect to S_t:

$$\frac{\partial \text{Vega}}{\partial S_t} \qquad (95)$$

This derivative is of interest to a vega trader. Similarly, the partial derivative of all important Greeks with respect to a small change in time parameter may provide information about the way the Greeks move over time.

6.7. Greeks and PDEs

The fundamental Black-Scholes POE that we derived in this chapter can be reinterpreted using the Greeks just defined. In fact, we can plug the Greeks into the black-Scholes PDE

$$\frac{1}{2} C_{SS}\sigma^2 S_t^2 + rC_S S_t - rC + C + C_t = 0 \qquad (96)$$

and recast it as

$$\frac{1}{2} \text{Gamma } \sigma^2 S_t^2 + r\text{Delta } S_t - rC + \text{Theta} = 0 \qquad (97)$$

In this interpretation, being long in options means, "earning" gamma and "paying" theta.

It is also worth noting that the higher order Greeks mentioned in equations (94) and (95) are not present in equation (97). This is because they are second order Greeks. The first order Greeks are related to changes in the underlying risk ΔS_t, Δs or time Δ, whereas the higher order Greeks would relate to changes that will have sizes given by the products ($\Delta S_t \; \Delta s$)or ($\Delta s \; \Delta$). In fact, when ΔS_t, Δs , Δ are "small" but non-negligible, products of two small numbers such as ($\Delta S_t \; \Delta s$) are even smaller and negligible, depending on the sizes of incremental changes in S_t. or volatility.[22]

In some real life applications, when volatility "spikes", higher order Greeks may become relevant. Yet, in theoretical models with standard assumptions, where $\Delta \to 0$, they fall from the overall picture, and do not contribute to the POE in equation (96).

6.7.1. Gamma Trading

The Black-Scholes POE can be used to explain what a gamma trader intends to accomplish. Assume that the real-life gamma is correctly calculated by choosing a formula for C(S_t, t r, K, s , T) and then taking the derivative:

$$\text{Gamma} = \frac{\partial^2 C(S_t, t \,|\, r, K, \sigma, T)}{\partial S_t^2} \qquad (98)$$

Following the logic that led to the Black-Scholes POE in equation (96), a gamma trader would, first, form a subjective view on the size of expected changes in the underlying using some subjective probability P*, as of time to < t. The gains can be written as,

$$\frac{1}{2}\text{Gamma}\left(E^{P^*}{}_{t0}[(\Delta S_t)^2]\right) \qquad (99)$$

This term would be greater, the greater the oscillations in S_t. Then, these gains will be compared with interest expenses and the loss of time value. If the expected gamma gains are greater than these costs, then the gamma trader will go long gamma. If, in contrast, the costs are greater, the gamma trader will prefer to be short gamma.

There are at least two important comments that need to be made about trading gammas.

6.7.2. Gamma Trading versus Vega

First of all, the gamma of an option position depends on the implied volatility parameter s . This parameter represents implied volatility. It need not have the same value as the (percentage) oscillations anticipated by a gamma trader. In fact, a gamma trader's subjective (expected) gains, due to S_t oscillations, are given by

$$\frac{1}{2} \text{Gamma} \quad E^{P^*}{}_{t0}[(\Delta S_t)^2] \tag{100}$$

There is no guarantee that the implied volatility parameter will satisfy the equality

$$\sigma^2 S^2_t \Delta = E^{P^*}{}_{t0}[(\Delta S_t)^2] \tag{101}$$

This is, even if the trader is correct in his or her anticipation. The right-hand side of this expression represents the anticipated (percentage) oscillations in the underlying asset that depend on a subjective probability distribution, whereas the left-hand side is the volatility value that is plugged into the Black-Scholes formula to get the option's fair price.

Thus, a gamma trader's gains and losses also depend on the implied volatility movements, and the option's vega will be a factor here. For example, a gamma trader may be right about increased, real-world oscillations, but, may still lose money if implied volatility, s , falls simultaneously. This will lower the value of the position if

$$\frac{\partial Css}{\partial s} < 0 \tag{102}$$

The following reading illustrates the approaches a trader or risk manager may adopt with, respect to vega and gamma risks.

Example

The VOLX contracts, (one) the new futures based on the price volatility of three reference markets measured by the closing levels of the benchmark cash index. The three are the German (DAX), UK (FT-SE), and Swedish (OMX) markets.

The designers argue that VOLX products, by creating a term structure of volatility that is arbitrageable, offer numerous hedging and trading possibilities. This covers both Vega and Gamma exposures and also takes in the long-dated options positions that are traditionally very difficult to hedge with short options.

Simply put, option managers who have net short positions and therefore are exposed to increases in volatility, can hedge those positions by being long the VOLX contract. The reverse is equally true. As a pure form of Vega, the contracts offer particular benefits for Vega hedging. Their Vega profile is constant for any level of spot ahead of the rare setting period, and then diminishes linearly once the RSP has begun.

The gamma of VOLX futures, in contrast, is very different from those of traditional options. Although a risk manager would traditionally hedge an option position by using a product with a similar gamma profile hedging the gamma of a complex book with diversified strikes can become unwieldy. VOLX gamma, regardless of time and the level underlying spot, is evenly distributed. VOLX will be particularly useful for the traditionally hard to hedge out-of-the-money wings of an option portfolio. (IFR, November 23, 1996).

6.7.3. Which Expectation?

We characterized trading gains expected from S_t-oscillations using the expression:

$$\frac{1}{2}\text{Gamma}\left(E^{P^*}{}_{t0}[(\Delta S_t)^2]\right) \tag{103}$$

Here the expectation $E^{P^*}{}_{t0}[(\Delta S_t)^2]$ is taken with respect to subjective probability distribution P*. The behavior of gamma traders depends on their subjective probability, but the market-determined arbitrage-free price will be objective and the corresponding expectation has to be arbitrage-free. The corresponding pricing formulas will depend on objective risk-neutral probabilities.

7. Real-Life Complications

In actual markets, the issues discussed here should be applied with care, because there will be significant deviations from the theoretical Black-Scholes world. By convention, traders consider the Black-Scholes world as the benchmark to use, although its shortcomings are well known.

Every assumption in the Black-Scholes world can be violated. Sometimes, these deviations are harmless or can easily be accommodated by modifying the formula. Some such modifications of the formula would be minor, and others more significant, but, in the end they take care of the problem at a reasonable effort.

Yet, there are two cases that require substantial modifications. The first concerns the behav-ior of volatility. In financial markets, not only is volatility not constant, but it also has some unexpected characteristics. One of these anomalies is the smile effects.[23] Volatility has, also, a term structure.

The second case is when interest rates are stochastic, and, the underlying asset is an interest--rate-related instrument. Here, the deviation from the Black-Scholes world, again, leads to significant changes.

7.1. Dealing with Option Books

This lesson discussed gamma, delta, and vega risks for the case of a single option position. Yet, market makers do not deal with single options. They have option books and they try to manage the delta, gamma, and vega risks of portfolios of options. This complicates the hedging and risk management significantly. The existence of exotic options compounds these difficulties.

First of all, option books consist of options on different, possibly correlated, assets. Second, implied volatility may be different across strikes and expiration dates, and a straightforward application of delta, gamma, and vega concepts to the portfolio may become impossible. Third, while for single options delta, vega, and gamma have known shapes and dynamics, for portfolios of options, the shapes of delta, gamma, and vega are more complex and their movement over time may be more difficult to track.

7.2. Futures as Underlying

This lesson has discussed options written on cash instruments. How would we analyze options that are written on a futures or forward contract? There are two steps in designing option contracts. First, a futures or a forward contract is introduced on the cash instrument, and second, an option is written on the futures. The holder of the option has the right to buy one, or more, futures contract.

Why would anyone write an option on futures (forwards), instead of writing it on the cash instrument directly?

In fact, the advantages of such contracts are many, and the fact that option contracts written on futures and forwards are the most liquid is not a coincidence. First of all, if ‘one were to buy and sell the underlying in order to hedge the ‘option positions, the futures contract are more convenient. They are more liquid, and they do not require upfront cash payments. Second, hedging with cash instruments would imply, for example, selling or buying thousands of barrels of oil. Where would a trader put so much oil, and where would he get it? Worse, dynamic hedging requires adjusting such positions continuously. It would be very inconvenient buy and sell cash underlying. Long and short positions in futures do not result in delivery until the expiration date. Hence, the trader can constantly adjust his or her position without having to store barrels of oil at each rebalancing of the hedge. Futures are also more liquid and the associated transactions costs and counterparty risks are much smaller.

Hence, the choice of futures and forwards as the underlying instead of cash instruments is, in fact, clever contract design. But, we must remember that futures come with daily marking to market. Forward contracts, on the other hand, may not require any marking to market until the expiration date.

7.2.1. Delivery Mismatch

Note the possibility of a mismatch. The option may result in the delivery of a futures contract at time T, but the futures contract may not expire at that same time. Instead, it may expire at a time $T + \Delta$ and may result in the delivery of the cash commodity. Such timing mismatches introduce new risks.

8. Conclusion: What Is an Option?

This lesson has shown that an option is essentially a volatility instrument. The critical parameter is how much the underlying risk oscillates within a given interval. We also saw that there are many other risks to manage. The implied volatility parameter, s, may change, interest rates may fluctuate, and option sensitivities may be have unexpectedly. These risks are not “costs” of maintaining the position perhaps, but they affect pricing and play an important role in option trading.

ENGINEERING CONVEXITY POSITION

Objectives

- This lesson explains you about the dependence of pricing equations on risk factors and how the different instruments are related to it.

How can anyone trade volatility? Stocks, yes, Bonds, yes. But volatility is not even an asset. Several difficulties are associated with defining precisely what volatility is. For example, from a technical point of view, should we define volatility in terms of the estimate of the conditional standard deviation of an asset price S_t?

$$\sqrt{E_t[S_t - E_t[S_t]]^2} \qquad (1)$$

Or should we define it as the average absolute deviation?

$$E_t[\,|S_t - E_t[S_t]|\,] \qquad (2)$$

There is no clear answer, and these two definitions of statistical volatility will yield different numerical values. Leaving statistical definitions of volatility aside, there are many, instances where traders quote, directly, the volatility instead of the dollar value of an instrument. For example, interest rate derivatives markets quote cap-floor and swaption volatilities. Equity options provide implied volatility. Traders and market makers trade the quoted volatility. Hence, there must be some way of isolating and pricing what these traders call volatility in their respective markets.

We started seeing how this can be done in previous lessons. Convexity of options became more valuable when "volatility" increased. In the previous lessons you have seen how these strategies can quantify and measure the "volatility" of an asset in monetary terms. This was done by forming delta-neutral portfolios, using assets with different degrees of convexity. In this lesson, we develop this idea further, apply it to instruments other than options, and obtain some generalizations. The plan for this lesson is as follows.

First we show how convexity of a long bond relates to yield volatility. The higher the volatility of the associated yield, the higher the benefit from holding the bond. We will discuss the mechanics of valuing this convexity. Then, we compare these mechanics with option-related convexity trades. We see some close similarities and some differences. At the end, we generalize the results to any instrument with different convexity characteristics.

A Puzzle

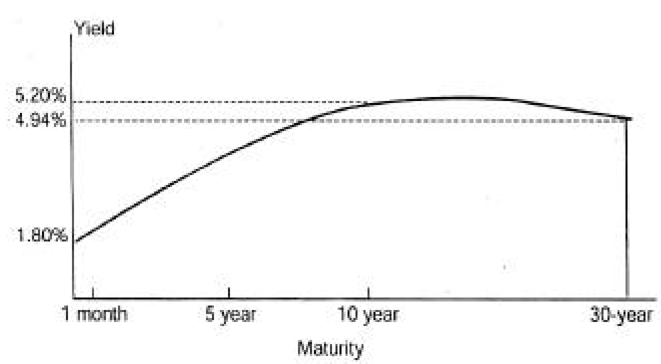
Here is a puzzle. Consider the (fictititious) yield curve shown in Figure 9-1. The to-year zero coupon bond has a yield to maturity that equals 5.2%: The- 30-year zero, however, has a yield to maturity of just 4.94%. In other words, if we buy and hold the latter bond 20 more years, we would receive a lower yield during its lifetime.

It seems a bit strange that the longer maturity is compensated with a lower yield. There are several economic or institutional explanations of this phenomenon. For example, expectations for inflation 20 years down the line may be less than the inflationary expectations for the next 10 years only. Or, the relative demands for these maturities may be determined by institutional factors and, because players don't like to move out of their preferred maturity, the yield curve may exhibit such inconsistencies. Insurance companies, for example, need to hedge their positions on long-term retirement contracts and this preference may lower the yield and raises the price of long bonds.

But, these explanations can hardly fully account for the observed anomaly. Institutional reasons such as preferred habitat and treasury debt retirement policies that reduce the supply of 30-year treasuries may account for some of the difference in yield, but it is hard to believe that additional 20-year duration is compensated so little. Can there be another explanation?

In fact, the yield to maturity may not show all the gains that can be realized from holding a long bond. This may be hard to believe, as yield to maturity is by definition how much the bond will yield per annum if kept until maturity.

Yet, there can be additional gains to holding a long bond, due to the convexity properties of the instrument, depending on what else is available to trade "against" it, and depending on the underlying volatility. These could explain the "puzzle" shown in Figure 9-1. The 4.94% paid by the 30-year treasury, plus some additional gains could exceed the total return from the to-year bond. This is conceivable since the yield to maturity and the total return of a bond are, in fact, quite different ways of measuring financial returns on fixed-income instruments.



Bond Convexity Trades

We have already seen convexity trades within the context of vanilla options. Straightforward discount bonds, especially those with long maturities, can be analyzed in a similar fashion and have exposure to interest rate volatility. In fact, a "long" bond and a vanilla option are both convex instruments and they both coexist with instruments that are either linear or have less convexity.[1] Hence, a delta-neutral portfolio can be put together for long maturity bonds to benefit from volatility shifts. The overall logic will be similar to the options discussed in the previous chapter.

Consider a long maturity default-free discount bond with price B (t, T), with t < T. This bond's price at time t can be expressed using the corresponding time t yield, y^T_t:

$$B(t,T) = \frac{1}{(1+y^T_t)^T} \tag{3}$$

Fort = 0, and T = 30, this function is plotted against various values of the 30-year yield, in Figure 9-2. It is obvious that the price is a convex function of the yield.

A short bond, on the other hand, can be represented in a similar space with an almost linear curve. For example, Figure 9-3 plots a 1-year bond price B(O, 1) against a 1-year yield y^1_0.We see that the relationship is essentially linear.[2]

The main point here is that, under some conditions, using these two bonds we can put together a portfolio that will isolate bond convexity gains similar to the convexity gains that the dynamic hedging of options has generated. Thus, suppose movements in the two yields y^1_t and y^{30}_t are perfectly correlated over time t.[3] Next, consider a trader who tries to duplicate the strategy of the option market maker-discussed in the previous chapter. The trader buys the long: bond with borrowed funds and delta-hedges the first-order yield exposure by shorting an appropriate amount of the shorter maturity bond.

This trader will have to borrow B(O, 30) dollars to buy and fund the long bond position. The payoff of the portfolio

{Long bond, loan of B(O, 30) dollars} (6)

is as shown in Figure 9-2b as curve BB'. Now compare this with the bottom part of Fig-ure 9-2. Here we show the profit/loss position of a market maker who buys an at-the-money "put option" on the yield y^{30}_t. At expiration time T, the option will pay

$$P(T) = \max[y^{30}_t - y^{30}_t, 0] \tag{7}$$

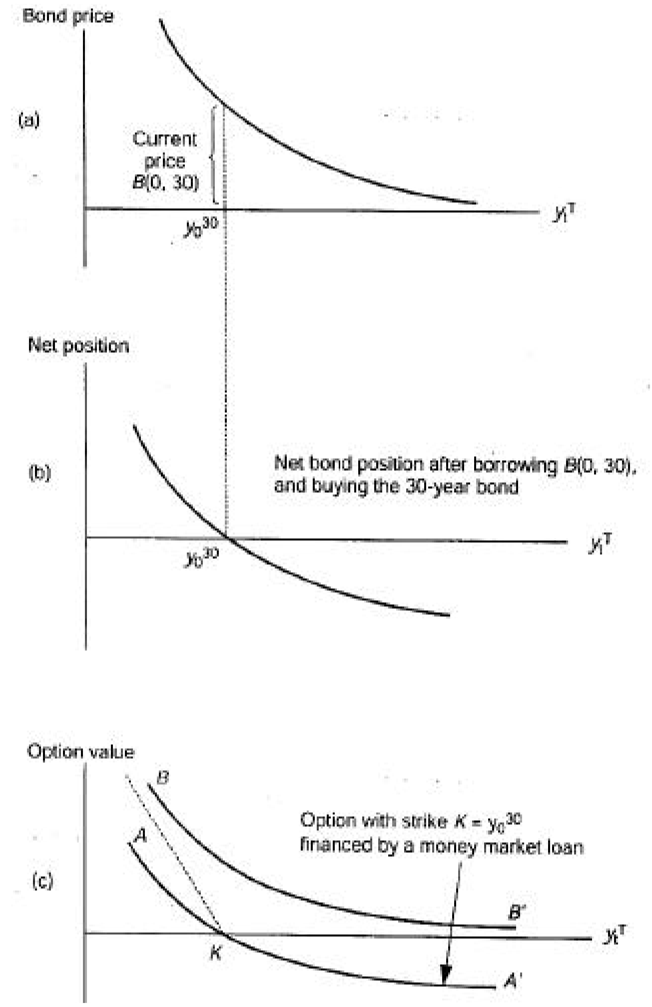
[1] The short maturity bonds are almost linear, In the case of vanilla options, positions on underlying assets such as' stocks are also linear. In fact, a first-order Taylor series expansion around zero yields

$$B(0,1) = \frac{1}{(1+y^1_0)} \tag{4}$$

$$\bullet(1+Y^1_0) \tag{5}$$

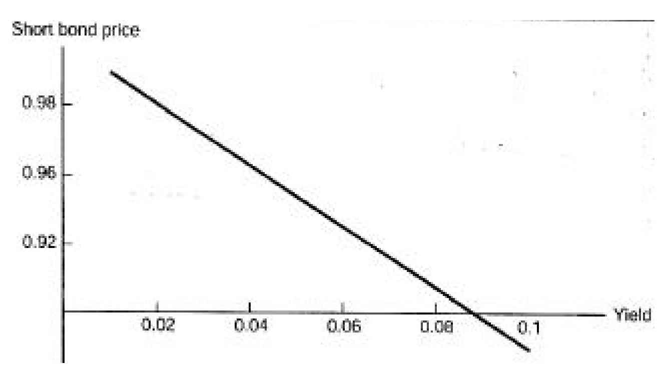
if the y^1_0 is "small,"

[3] This simplifying assumption implies that all bonds are affected by the same unpredictable random shock, albeit to a varying degree. It is referred to as the one factor model.



This option is financed by a money market loan so that the overall position is shown as the downward sloping curve AA'.[4] We see a great deal of resemblance between the two positions. Given this similarity between bonds and options, we should be able to isolate convexity or gamma trading gains in the case of bonds as well. In fact, once this is done, using an arbitrage

[4] The option price is the curve BB'. The curve shifts down by the money market loan amount P_o, which makes the position one of zero cost.



argument, we should be able to obtain a partial differential equation (PDE) that default-free discount bond prices will satisfy. This PDE will have close similarities to the Black-Scholes PDE derived in Chapter 8.

The discussion below proceeds under some simplifying and unrealistic assumptions. We use the so-called one-factor model. Our purpose is to understand the mechanics of volatility trading in the case of bonds and this assumption simplifies the exposition significantly. Our contexts different than in real life, where fixed-income instruments are affected by more than a single common random factor. Thus, we make two initial assumptions:

1. There is a short and a long default-free discount bond with maturities T s and T, respectively. Both bonds are liquid and can be traded without any transaction costs.
2. The two bond prices depend on the same risk factor denoted by T_t. This can be interpreted as a spot interest rate that captures all the randomness at time t, and is the single factor mentioned earlier.

The second assumption means that the two bond prices arc a function of the short rate. r_t. These functions can be written as

$$B(t, T^S) = S(r_t, t, T^S) \tag{8}$$

$$B(t., T) = B(r_t, t, T) \tag{9}$$

where $B(t, T^S)$ is the time-t price of the short bond and the B(t, T) is the time-t price of the long bond We postulate that the maturity T^S is such that the short bond price $B(t, T^s)$ is (almost)a lin-ear function of r_t. meaning that the second derivative of $B(t, T^S)$ with respect to r_t is negligible.

Thus, we will proceed as if there was a single underlying risk that causes price fluctuations in a convex and a quasi-linear instrument, respectively. We will discuss the cash gains generated by the dynamically-hedged bond portfolio in this environment.

3.1. Delta/Hedged Bond Portfolios

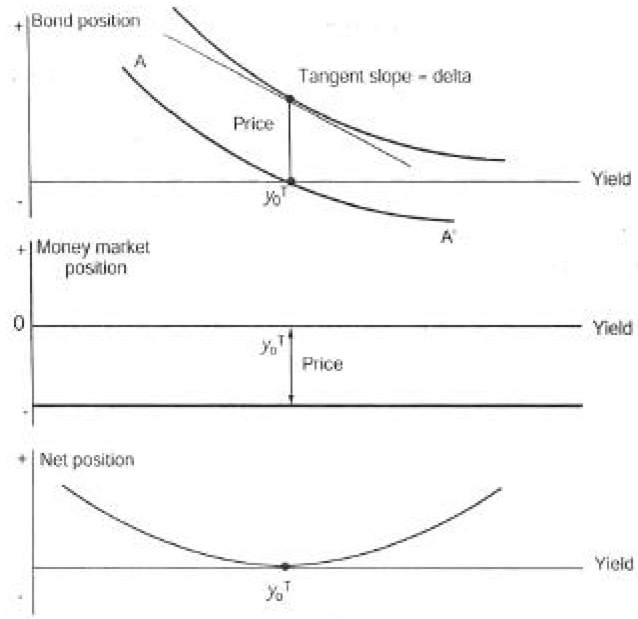
The trader buys the long bond with borrowed funds and then hedges the downside risk implied by the curve AA' in Figure 9-4. The hedge for the downside risk will be a position that makes money when r_t increases, and loses money when r_t declines. This can be accomplished by shorting an appropriate number of the short bond.

In fact, the trick to form a delta-neutral portfolio is the same as in Chapter 8. Take the partial derivative of the functions $S(r_t, t, T^s)$ and $B(r_t, t, T)$ with respect to r_t, evaluate them at point $r_{to,}$ and use these to form a hedge ratio, h_t:

$$h_t = \frac{\left(\frac{\partial B(r_t, t, T)}{\partial r_t}\right)}{\left(\frac{\partial S(r_t, t, T^S)}{\partial r_t}\right)} \tag{10}$$

$$= \frac{B_r}{S_r} \tag{11}$$

The S_r is assumed to be a constant, given the quasi-linearity of the short bond price with respect to r_t. The h_t is a function of r_t. since the B_r is not constant due to the long bond's convexity. Given the value of r_{to}' the h_t can be numerically calculated, and h_{to} units of the short maturity bond would be sold short at t_o.



The change in the value of this portfolio due to a small change in the spot rate Δr_t only, is given by

$$\Delta [B(r_t, t, T) - h_t S(r_t, t, T^S)] = B_r \Delta r_t - \frac{B_r}{S_r} S_r \Delta r_t + R$$
$$= R$$

since the S_r terms cancel out. R is the remainder term of the implied Taylor series approximation, or Ito's Lemma, which depends essentially on the second derivative, B_{rr}' and on r_t volatility. The S_r is approximately constant. This means that the net position,

{ Borrow B(t, T) dollars, Buy one B(t, T), Short ht units of $B(t, T^S)$ } (14)

will have the familiar volatility position shown in the bottom part of Figure 9-4. As r_t fluctuates, this position is adjusted by buying and (short) selling an appropriate number of the nonconvex asset The new value of partial derivative, h_t, is used at each readjustment. Again, just like: in , Chapter 8, this will make the practitioner "sell high" and "buy low" (or vice versa). As a result of these hedge-adjustments, the counter-party who owns the long bond will earn gamma profits. These trading gains will be greater as volatility increases. Hence, we reach the result:

- Everything else being the same, the greater the volatility of rt. the more "valuable" will the long bond be.

This means that as volatility increases, ceteris paribus, the yield of the convex instruments should decline, since more market participants will try to put this trade in place, and drive its price higher.

Example

Suppose that initially the yield curve is flat at 5%. The value of a 30-year default-free discount bond is given by

$$B(O,30) = \frac{1}{(1+.05)^{30}} \tag{15}$$

$$= \quad 0.23 \tag{16}$$

The original delta of the bond, D_{to} at r_{to} = .05 will be:

$$D_{to} = \frac{30}{(1+r_{to})^{31}} \tag{17}$$

$$= \quad 6.61 \tag{18}$$

A I-year short bond is assumed to have an approximately linear pricing formula

$$B(t_o, T^s) = (1 - r_{to}) \tag{19}$$

$$= \quad 0.95 \tag{20}$$

The market maker will borrow 0.23 dollars, buy one long bond, and then hedge this position by shorting

$$\frac{-6.61}{-1.0} \tag{21}$$

units of the short bond. (Given linearity approximation short bond has unit interest sensitivity.)

A small time, Δ, passes. All rates change. r_t moves to 6%. The portfolio value will move

$$\Delta B(t,T) - h_t \Delta B(t,T^s) = \left[\frac{1}{(1+.06)^{30}} - \frac{1}{(1+.05)^{30}}\right] - 6.61\,[(1-.06)-(1-.05)] \tag{22}$$

$$= 0.009 \tag{23}$$

Note that in calculating this number, we are assuming that " is small. Only the effect of changing r_t is taken into account. /n a sense, we are using a framework similar to partial derivatives.

The new delta is calculated as -4.9. The adjusted portfolio should be short 4.9 units of the short bond. Thus,

$$(6.6 - 4.9) = 1.7 \tag{24}$$

units need to be covered at a price of 0.94 each to bring the position to the desired delta-neutral state.

This leaves a trading profit equal to

$$1.7(0.95 - 0.94) = .017\$ \tag{25}$$

Another period passes, with rt going back to r_{t2} = .05. The cycle repeats itself. The delta will change again, the portfolio will be readjusted, and trading profits will continue to accumulate.

This example is approximate, since not all costs of the position are taken into account. The example started with the assumption of a flat yield curve, which was later relaxed and the yields became volatile. However, we never mentioned what causes this change. It turns out that, volatility leads to additional gains for long bond holders, and this increases the demand for them. As a result, ceteris paribus, long bond yields would decline relative to short bond yields. Hence, the introduction of yield volatility changed the structure of the initial yield curve.

3.2. Costs

What are the costs (and other gains) of putting together such a long volatility position using default-free discount bonds? First, there is the funding cost. To buy the long bond, B (t, T) funds were rowed at r_t percent per annum. As long as the position is kept open, interest expense will be incurred. Second, as time passes, the pricing function of the bond becomes less and less convex, and hence the portfolio's trading gains will respond less to volatility changes. Finally, as time passes, the value of the bonds will increase automatically even if the rates don't come down.

3.3. A Bond PDE

A partial differential equation can be put together consisting of the gains from convexity of long. bonds, and costs of maintaining the volatility position. Under some conditions, this PDE has, an analytical solution, and an analytical formula can be obtained the way the Black-Scholes formula was obtained.

First we discuss the PDE informally. We start with the trading gains due to convexity. These gains are given by the continuous adjustment of the hedge ratio ht, which essentially depends on the B_r, except for a constant of proportionality, since the hedging instrument is quasi-linear in r_t. As r_t changes, the partial B_r changes, and this will be captured by the second derivative. Then, convexity gains during a small time interval Δ. is a function, as in Chapter 8, of

$$\frac{1}{2}\frac{\partial^2 B(t,T)}{\partial r_t^2}\left(\sigma(r_t, t)\, r_t\sqrt{\Delta}\right)^2 \tag{26}$$

This is quite similar to the case of vanilla options, except that here the s (r_t, t) is the percentage short rate volatility. Short bond interest sensitivity will cancel out.

If we model the risk-neutral dynamics of the short rate r_t as

$$d_{rt} = \mu(r_t, t)dt + s r_t dW_t\, t \cdot [O,T] \tag{27}$$

where percentage volatility s is constant, these gamma gains simplify to

$$\frac{1}{2} B_{rr}\, \sigma^2 r_t^2 \Delta \tag{28}$$

during a small period Δ.[5]

To these, we need to add (subtract) other costs and gains that the position holder is subject to. The interest paid during the period Δ. on horrowed funds will be

$$r_t B(t,T)\Delta \tag{30}$$

The other gain (loss) is the direct effect of passing time

$$\frac{\partial B(t,T)}{\partial t}\Delta = B_t\Delta \tag{31}$$

As time passes, bonds earn accrued interest, and convexity declines. The interest earned due to shorting the linear instrument will cancel out the cost of this short position.

The final component of the gains and losses that the position is subject to during Δ is more complex than the case of a vanilla call or put. In the case of the option, the underlying stock, S_t , provided a very good delta-hedging tool. The market maker sold $C(S_t,t)$ units of the $\frac{\partial}{\partial S_t}$

[5] Note that we are using the notation

$$\frac{\partial^2 B(t,T)}{\partial r_t^2} = B_{rr} \qquad (29)$$

underlying S_t in order to hedge .1 long call position. In the present case, the underlying risk is not the stock price S_t or some futures contract. The underlying risk is the spot rate r_t, and this is not an asset. That is to say, the "hedge" is not r_t itself, but instead it is an asset indirectly influenced by r_t. Also, randomness of interest rates requires projecting future interest gains and costs. All these complicates the cash flow analysis.

These complication can be handled by positing that the drift term μ (rt, t) in the dynamics, [6]

$$d_{rt} = \mu(r_t, t)dt + s r_t dWt, t \bullet [O,T] \qquad (32)$$

represents the risk-free expected change in the spot rate over an infinitesimal interval dt.[7] Using this drift, we can write the last piece of gains and losses over a small interval Δ as (Vasicek (1977))

$$\mu(r_t, t)B_r\Delta \qquad (33)$$

Adding all gains and losses during the interval Δ, we obtain the net gains from the convexity position:

$$\frac{1}{2} B_{rr}\sigma^2 r^2_t \Delta + \mu(r_t,t) B_r\Delta - r_t B\Delta + B_t\Delta \qquad (34)$$

In order to preclude arbitrage opportunities, this sum must equal zero. Cancelling the common Δ terms, we get the PDE for the bond:

$$\frac{1}{2} B_{rr}\sigma^2 r^2_t + \mu(r_t,t) B_r - r_t B + B_t = 0 \qquad (35)$$

The boundary condition is simpler than in the case of vanilla options and is given by

$$B(T,T) = 1, \qquad (36)$$

the par value of the default-free bond at maturity date T.

3.4. PDEs and Conditional Expectations

In this PDE, the unknown is again a function B(t,T). This function will depend on the random process r_t , the t, as well as other parameters of the model. The most important of these is the short rate volatility, s . If r_t is the continuously compounded short rate, the solution is given by the conditional expectation

$$B(t,T) = E^{P}_t\left(e^{-\int_t^T r_u du} \right) \qquad (37)$$

where, P is an appropriate probability. In other words, taking appropriate partial derivatives of the right-hand side of this expression, and then plugging these in the PDE would make the sum; on the left-hand side of equation (35) equal to zero.[8]

[6] See the Appendix to Chapter 8 for a definition of this SDE.

[7] Chapter 11 will go into the details of this argument that uses risk-neutral probabilities.

[8] The major condition to be satisfied for this is the Markovness of r_t.

It is interesting to look at the parallel with options. The pricing function for B (t, T) was based on a particular conditional expectation and solved the bond PDE. In the case of vanilla options written on a stock S_t, and satisfying all Black-Scholes assumptions, the call price $C(S_t, t)$ is given by a similar conditional expectation,

$$C(S_t, t) = E^P_t [e^{-r(T-t)} C(S_T, T)] \qquad (38)$$

where T is the expiration date, and, P is the appropriate probability. If this expectation is differentiated with respect to S_t and t, the resulting partial derivatives will satisfy the Black--Scholes PDE with the corresponding boundary condition. The main difference is that the Black- - Scholes assumptions take the short rate r_t to be constant, whereas in the case of bonds, it is a, stochastic process.

These comments reconcile the two views of options that were mentioned in Chapter 8.lf we interpret options as directional instruments, then equation (38) will give the expected gains of the optional at expiration, under an appropriate probability. The argument above shows, that, this expectation solves the associated PDE which was approached as an arbitrage relationship tying Gamma gains to other costs incurred during periodic rebalancing. In fact, we see that the two interpretations of options are equivalent.

3.5. From Black-Scholes to Bond POE

Comparing the results of trading bond convexity with those obtained in trading vanilla options provides good insights into the general characteristics of PDE methods that are commonly used in finance.

In Chapter 8, we derived a PDE for a plain vanilla call, C (t) using the argument of convexity trading. In this chapter, we discussed a PDE that is satisfied by a default-free pure discount bond B(t, T). The results were as follows.

1. The price of a plain vanilla call, written on a nondividend paying stock S_t, strike K, expiration T, was shown to satisfy the following "arbitrage" equality

$$\frac{1}{2} C_{ss} (\sigma(S_t,t)S_t)^2 \Delta = (rC - rC_s S_t)\Delta - C_t \Delta \qquad (39)$$

wheres (S_t, t) is the percentage volatility of S_t during one year. The way it is written here, this percentage volatility could very well depend on time t, and S_t.

According to this equation, in order to preclude any arbitrage opportunities, trading gains obtained from dynamic hedging during a period of length Δ should equal the net. ' funding cost, plus loss of time value. Canceling common terms and introducing the' boundary condition yielded the Black-Scholes PDE for a vanilla call:

$$½ C_{ss} (\sigma(S_{t,t})S_t)^2 + rC_sS_t - rC + C_t = 0 \qquad (40)$$

$$C(T) = \max[ST - K, 0] \qquad (41)$$

Under the additional assumption that s $(S_t, t)S_t$ is proportional to S_t with a constant factor of proportionality s

$$\sigma(S_t, t)\, S_t = \sigma S_t \quad (41)$$

this PDE could be solved analytically, and a closed-form formula could be obtained for the C(t). This formula is the Black-Scholes equation:

$$C(t) = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \quad (42)$$

$$d_{1,2} = \frac{\log^{S_t/K} + r(T,t) \pm \frac{1}{2}\sigma^2 (T-t)}{\sigma\sqrt{T-t}} \quad (44)$$

The partial derivatives of this C(t) would satisfy the preceding PDE.

2. The procedure for a default-free pure discount bond B(t, T) followed similar steps, with some noteworthy differences, Assuming that the continuously compounded spot interest rate, r_t, is the only factor in determining bond prices, the convexity gains due to oscillations in r_t and to dynamic hedging can be isolated, and a similar "arbitrage relation" can be obtained:

$$\frac{1}{2} B_{rr} (\sigma(r_t,t) r_t)^2 \Delta = (r_t B - \mu(r_t,t) B_r) \Delta - B_t \Delta \quad (45)$$

Here, the s (r_t, t) is the percentage volatility of the short rate r_t during one year.

Cancelling common terms, and adding the boundary condition, we obtain the bond PDE:

$$\frac{1}{2} B_{rr} \sigma^2 r_t^2 + \mu(r_t,t) B_r - r_t B + B_t = 0 \quad (46)$$

$$B(T,T) = 1 \quad (47)$$

Under some special assumptions on the dynamic behavior of r_t, this bond PDE can be solved analytically, and a closed-form formula can be obtained.

We now summarize some important differences between these parallel procedures. First, note that the PDE for the vanilla option is obtained in an environment where the only risk comes from the asset price S_t, whereas for bonds the only risk is the interest rate r_t, which is not an asset per se. Second, the previously mentioned difference accounts for the emergence of the term $\mu(r_t, t)$ in the bond PDE, while no such non-transparent term existed in the call option PDE. The $\mu(r_t, t)$ represents the expected change in the spot rate during dt once the effect of interest rate risk is taken out. Third, the $\mu(r_t, t)$ may itself depend on other parameters that affect interest rate dynamics. It is obvious that under these conditions, the closed-form solution for B(t, T) would depend on the same parameters. Note that in the case of the vanilla option, there was no such issue and the only relevant parameter was a, This point is important since it could make the bond price formula depend on all the parameters of the underlying random process, whereas in the case of vanilla options, the Black-Scholes formula depended on the characteristics of the volatility parameter only.

Before we close this section, a final parallel between the vanilla option and bond prices should be discussed. The PDE for a call option led to the closed-form Black-Scholes formula under some assumptions concerning the volatility of S_t. Are there similar closed-form solutions to the bond PDE? The answer is yes.

3.6. Closed -Form Bond Pricing Formulas

Under different assumptions concerning short rate dynamics, we can indeed solve the bond PDE and obtain closed-form formulas, We consider three cases of increasing complexity. The cases are differentiated by the assumed short rate dynamics.

3.6.1. Case 1

The first case is simple. Suppose rt is constant at r. This gives the trivial dynamics,

$$dr_t = 0 \quad (48)$$

where a and $\mu(r_t, t)$ are both zero. The bond PDE in equation (46) then reduces to

$$-rB + Bt = 0 \quad (49)$$

$$B(T,T) = 1 \quad (50)$$

This is a simple ordinary differential equation. The solution B(t, T) is given by

$$B(t, T) = e^{-r(T-t)} \quad (51)$$

3.6.2. Case 2

The second case is known as the Vasicek model.[9] Suppose the risk-adjusted dynamics of the spot rate follows the mean-reverting process given by[10]

$$d_{rt} = a(k - r_t)dt + s\,dW_t,\ t \bullet [0,T] \quad (52)$$

where the W_t is a Wiener process defined for a risk-adjusted probability. [11]

Note that the volatility structure is restricted to constant absolute volatility denoted by s. Suppose, further, that the parameters a, k, s, are known exactly. The fundamental PDE for a typical B(t, T) will then reduce to

$$B_r a (k, - r_t) + B_t + \frac{1}{2} B_{rr}{}^{\sigma 2} - r_t B = 0 \quad (53)$$

Using the boundary condition B(T, T) = 1, this PDE can be solved analytically, to provide a closed-form formula for B(t, T). The closed-form solution is given by the expression

$$B(t, T) = A(t, T) e^{-C(t,T) rt} \quad (54)$$

where,

$$C(t,T) = \frac{(1 - e^{-a(T-t)})}{a} \quad (55)$$

9 See Vasicek (1977).

10 The fact that this dynamic is risk-adjusted is not trivial. Such dynamics are calculated under risk-neutral proba-bilities and may differ significantly from real-world dynamics. These issues will be discussed in Chapter 11.

3.6.3. Case 3

The third well-known case, where the bond POE in Equation (46) can be solved for a closed form, is the Cox-Ingersoll-Ross (CIR) model. In the CIR model, the spot rate r_t is assumed to obey the slightly different mean-reverting stochastic differential equation

$$dr_t = a(k - r_t)dt + \sigma\sqrt{r_t}\,dW_t \quad t \,\square\, [0,T] \quad (57)$$

which is known as the square-root specification of interest rate volatility. Here the Wt is a Wiener process under the risk-neutral probability.

The closed-form bond pricing equation here is somewhat more complex than in the Vasicek model. It is given by

$$B(t, T) = A(t, T)e^{-C(t,T)rt} \quad (58)$$

where the functions A(t, T) and C(t, T) are given by

$$A(t,T) = 2 \left(\frac{?^{e\ ½\ (a\ +?)\ (T-t)}}{(a + ?)\ (e^{?\ (T-t)} - 1) + 2?} \right)$$

and where ? is defined as

$$A(t,T) = 2 \quad \frac{?^{e\ ½\ (a\ +?)\ (T-t)}}{(a + ?)\ (e^{?\ (T-t)} - 1) + 2?} \quad (61)$$

The bond volatility s determines the risk premia in expected discount bond returns.

3.7 A Generalization

The previous sections showed that whenever two instruments depending on the same risk factor display different degrees of convexity, we can, in principle, put together a delta hedging strategy similar to the delta hedging of options discussed in Chapter 8. Whether this is worthwhile depends, of course, on the level of volatility relative to transactions costs and bid-ask spreads.

When a market practitioner buys a convex instrument and short-sells an appropriate number' of a linear (or less convex) instrument, he or she will benefit from higher volatility. We then say that the position is long volatility or long gamma. This trader has purchased gamma. If, in contrast, the convex instrument is shorted and the linear instrument is purchased at proper ratios, the position will benefit when the volatility of the underlying decreases.

As the case of long bonds shows, the idea that volatility can be isolated (to some degree), and then traded is very general, and can be implemented when instruments of different convexities are available on the same risk. Of course, volatility can be such that transaction costs and bid- ask spreads make trading it unfeasible, but that is a different point. M9re importantly, if the yield curve slope changes due to the existence of a second factor, the approach presented in the previous sections will not guarantee convexity gains.

4. Sources of Convexity

There is more than one reason for the convexity of pricing functions. We discuss some simple cases briefly, using a broad definition of convexity.

4.1. Mark to Market

We start with a minor case due to daily marking-to-market requirements. Let f_t denote the daily futures settlement price written on an underlying asset S_t let F_t be the corresponding forward price, and let r_t be the overnight interest rate. .

Marking to market means that the futures position makes or loses money every day depending on how much the futures settlement price has changed,

$$\Delta f_t = f_t - f_{t-1} \quad (62)$$

where the time index t is measured in days and hence is discrete.

Suppose the overnight interest rate T_t is stochastic. Then if the trader receives (pays) mark-to-market gains daily, these can be deposited or borrowed at higher or lower overnight interest rates. If Δf_t were uncorrelated with interest rate changes,

$$\Delta r_t = r_t - r_{t-1} \quad 63)$$

marking to market would not make a difference.

But, when St is itself an interest rate product or an asset price related to interest rates" ' the random variables Δf_t and Δr_t will, in general, be correlated. For illustration; suppose the . correlation between Δf_t and Δr_t is positive. Then, when f_t increases, r_t is likely to increase also, which means that the mark-to-market gains can now be invested at a higher overnight interest rate. If the correlation between Δf_t and Δr_t is negative, the reverse will be true. Forward contracts do not, normally, require such daily marking-to-market. The contract settles only at the expiration date. This means that daily paper gains or losses on forward contracts cannot be reinvested or borrowed at higher or lower rates.

Thus, a futures contract written on an asset St whose price is negatively correlated with T_t, will be cheaper than the corresponding forward contract. If the correlation between S_t and T_t is positive, then the futures contract will be more expensive. If S_t and T_t are uncorrelated, then futures and forward contracts will have the same price, everything else being the same.

Example

Consider any Eurocurrency future. We saw in Chapter 4 that the price of a J -year Eurodollar future, settling at time t + 1, is given by the linear function

$$V_t = 100(1 - f_t) \quad (64)$$

Normally, we expect overnight interest rate T_t to be positively correlated with the futures rate It. Hence, the price V_t, which is not a convex function, would be negatively correlated with T_t. This means that the Eurodollar futures will be somewhat cheaper that a corresponding forward contract, which in turn means that futures interest .rates are somewhat higher than the forward rates.

Mark-to-market is .one reason why futures and forward rates may be different.

4.2. Convexity by Design

Some products have convexity by design. The contract specifies payoffs and underlying risks, and this specification may make the contract price a nonlinear function of the underlying risks. Among the most important classes of instruments that permit such convexity gains arc, of course, options.

We also discussed convexity gains from bonds. Long maturity default-free discount bond prices, when expressed as a function of yield to maturity Yt, are simple nonlinear functions, such as

$$B(t,T) = \frac{100}{(1+yt)^T} \qquad (65)$$

Coupon bond prices can be expressed using similar discrete time yield to maturity. The price of a coupon bond with coupon rate c, and maturity T, can be written as .

$$P(t,T) = \sum_{i=1}^{T} \frac{100c}{(1+yt)^i} + \frac{100}{(1+yt)^T} \qquad (66)$$

It can be shown that default-free pure discount bonds, or strips, have more convexity than coupon . bonds with the same maturity.

4.2.1. Swaps

Consider a plain vanilla, fixed-payer interest rate swap with immediate start date at $t = t_0$ and end date, $t_n = T$. Following market convention, the floating rate set at time t_i is paid at time t_{i+1} For simplicity, suppose the floating rate is 12-month USD Libor. This means that $d = 1$. Let the time $t = t_0$ swap rate be denoted by s and the notional amount, N, be l.

Then, the time-to value of the swap is given by

$$V_{t_0} = E^P_{t_0}\left(\frac{L_{t_0} - s}{(1 + L_{t_0})} + \frac{L_{t_1} - s}{(1 + L_{t_0})(1 + L_{t_1})} + \ldots + \frac{L_{t_{n-1}} - s}{\prod_{i=0}^{n-1}(1 + L_{t_i})} \right) \qquad (67)$$

where $\{Lt_0, \ldots, Lt_{n-1}$ are random Libor rates to be observed at times $t_0, \ldots, t_{n-1}$,respec-tively, and P is an appropriate probability measure. With a proper choice of measure, we can act as if we can substitute a forward Libor rate, $F(t_0, t_i)$, for the future spot Libor L_{t_i} for all t_i [12]. If liquid markets exist where such forward Libor rates can be observed, then after this substitution, we can write the previous pricing formula as

$$V_{t_0} = \frac{L_{t_0} - s}{(1 + L_{t_0})} + \frac{F(t_0, t_1) - s}{(1 + L_{t_0})(1 + F(t_0, t_1))} + \frac{F(t_0, t_2) - s}{(1 + L_{t_0})(1 + F(t_0, t_1))(1 + F(t_0, t_2))} + \ldots + \frac{F(t_0, t_{n-1}) - s}{\prod_{i=0}^{n-1}(1 + F(t_0, t_i))} \qquad (68)$$

where $F(t_0, t_0) = L_{t_0}$, by definition. Clearly, this formula is nonlinear in each $F(t_0, t_i)$. As the forward rates change, the V_{t_0} changes in a nonlinear way.

This can be seen better if we assume that the yield curve is flat and that all yield curve shifts, are parallel. Under such unrealistic conditions, we have

$$L_{t_0} = F(t_0, t_0) = F(t_0, t_1) = \ldots\ldots = F(t_0, t_{n-1}) = F_{t_0} \qquad (69)$$

The swap formula then becomes

$$V_{t_0} = \frac{F_{t_0} - s}{(1 + F_{t_0})} \quad \frac{F_{t_0} - s}{(1 + F_{t_0})^2} \quad \frac{F_{t_0} - s}{(1 + F_{t_0})^T} \qquad (70)$$

which simplifies tol3

$$V_{t_0} = (F_{t_0} - s)\,\frac{((1 + F_{t_0})^T - 1)}{F_{t_0}(1 + F_{t_0})^T} \qquad (72)$$

The second derivative of this expression with respect to F_{t_0} will be negative, for all $F_{t_0} > 0$.

As this special case indicates, the fixed-payer swap is a nonlinear instrument in the underlying forward rates. Its second deriva-tive is negative, and the function is concave with respect: to a "typical" forward rate. This is not surprising, since, a fixed-payer swap has risks similar to the issuing of a 30-year bond. This means that a fixed-receiver swap will have a convex pricing formula and will have a similar profile as a long position in a 30-year coupon bond.

Example

Figure 9-5 plots the value of a fixed-payer swap under the restrictive assumption that the yield curve is flat and that it shifts only parallel to itself. The parameters are as follows.

$$t = 0 \qquad (73)$$

$$s = 7\% \qquad (74)$$

$$T = 30 \qquad (75)$$

$$F_{t0} = .06\% \qquad (76)$$

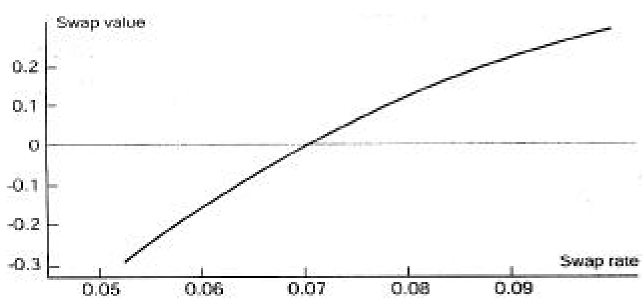
We see that the function is nonlinear and concave.

In Chapter 15 we will consider a different type of swap, with constant maturity. The convexity of constant maturity swaps is due to their structure. This convexity will, in general, be more' pronounced and at the same time more difficult to correctly account for.

Taking convexity characteristics of financial instruments into account is important. This is best illustrated by the Chicago Board of Trade's (CBOT) attempt to launch a new contract with 'proper convexity characteristics.

Example

The Chicago Board of Trade's board of directors last week approved a plan to launch 5- and l0 year U.S. dollar denomi-nated interest rate swap futures and options contracts.



Compared with the over-tile-counter market trading of swaps futures will reduce administrative cost and eliminate counter party risk, the exchange said.

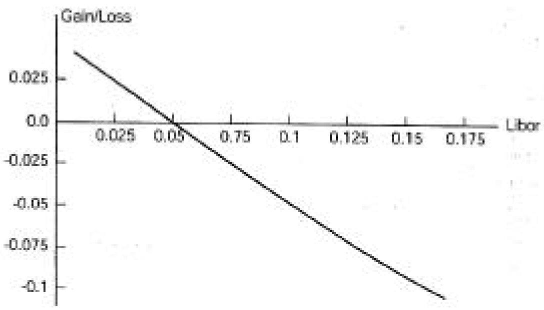
The CBOT's move marks the second attempt by the exchange to launch a successful swap futures contract. Treasuries were the undisputed benchmark a decade ago. They are not treated as a benchmark for valuation anymore: People price off the swap curve instead, said a senior economist at the CBOT.

The main difference between the new contract an4 the contract that the CBOT de-listed in the mid-1990s is that the new one is convex inform rather than linear. It's one less thing for end users to worry about, the economist said, noting that swap positions are marked to market on a convex basis. Another critical flaw in the old contract was that it launched in the three and five-year, rather than the five and ten-year maturities, which' is where most business takes place.

The new swaps contracts will offer institutional investors such as bank treasurers, mort-gage pass through traders, originators, service managers, portfolio managers, and other OTC market participants a vehicle for hedging credit and interest rate exposure, the exchange said. (IFR, Issue 1393, July 21, 2001)

This is an excellent example that shows the importance of convexity in contract design. Futures contracts are used for hedging by traders. If the convexity of the hedging instrument is different than the convexity of the risks to be hedged, then the hedge may deteriorate as volatility changes. In fact, as volatility increases, the more convex instrument may yield higher gamma gains and this will influence its price.

4.2.2. Convexity of FRAs



Now consider the case of forward rate agreements (FRAs). As discussed in Chapter 4, FRAs are instruments that can be used to fix, at time t_0, the risk associated with a Libor rate L_{ti} that will be observed at time t_i, and that has a tenor of ä expressed in days per year. [14] The question is when would this FRA be settled. This can be done in different ways, leading to slightly different instruments. We can envisage three types of FRAs.

One way is to set L_t, at time t_i, but then, settle at time $t_i + d$. This would correspond to the natural" way interest is paid in financial markets. Hence, at time t = to, the value of the FRA will be zero and at time $t_i + d$ the FRA buyer will receive or pay

$$[L_{ti} - F_{to}]N\,d \qquad (77)$$

depending on the sign of the difference. The FRA seller will have the opposite cash flow.

The second type of FRA trades much more frequently in financial markets. The description of these is the same, except that the FRA is settled at time t_i instead of at $t_i + d$. At time t_i when the Libor rate L_t. is observed, the buyer of the FRA will make (receive) the payment

$$\frac{[L_{ti} - F_{to}]\,d}{1 + L_{ti}\,d} N$$

This is the previous settlement amount discounted from time $t_i + d$ to time t_i, using the time t_i Libor rate. Figure 9-6 shows an example to the payoff of a 12 month FRA.

Of even more interest for our purpose is a third type of FRA contract, a Libor-in-arrears FRA, where the Libor observed at time t_i is used to settle the contract at time t_i according to

$$[L_{t\,i} - f_{to}]dN \qquad (79)$$

Here, f_{to} is the FRA rate that applies to this particular type of FRA.[15] Note that we are using a symbol different than the F_{to}, because the two FRA rates are, in general, different from each other due to convexity differences in the two contracts.

The question to ask here is under what conditions would the rates F_{to} and f_{to} differ from each other? The answer depends on the convexity characteristics of the underlying Contracts. In fact, market practitioners approximate these differences using convexity adjust-ment factors.

4.2.3. FRA Convexity Adjustments

As mentioned earlier, there are three types of FRAs. The rather non-liquid Libor-in-arrears FRAs make a payment at time t_i of the cash flow

$$Vt'i = [L_{ti}\ \ f_{to}]dN \qquad (80)$$

where d is days adjustment, and N is the notional amount as usual. The f_{to} is the forward rate associated with this Libor-in-arrears FRA. Note that the valuation fromula is linear in the interest rate observed at time t_i.

In market-traded FRAs, time-t_i settlement cash flows are discounted and are instead given by

$$V^*_{ti} = \frac{[L_{ti} - F_{to}]\,d}{(1 + L_{ti}\,d)} N$$

Here, the F_{to} is the forward rate discussed in Chapter 4. Note that the interest rate observed at time t_i is used to make a payment at the same time, after discounting a cash flow that was to be , received at time t_i+1. The formula is non-linear in this interest rate.

The two forward rates, F_{to} and f_{to}., cannot be identical. We can follow the reasoning that we introduced in Chapter 8. The two payoffs in equations (HO) and (81) have different convexity characteristics, if the two FRA rates were the same, a market practitioner could buy one FRA while selling the other in the right proportions and may end up with extra (convexity) gains.[16] Thus, the quoted rates on the two types of FRAs will have to be slightly different to compensate for such convexity differ-ences.

4.2.4. Swap Rate Adjustments

Plain vanilla swaps are convex instruments and are paid in arrears. There are also the so-called Libor-in-arrear swaps that use the time t_i Libor rate for the settlement at time t_i. Forward swap rates from this and from plain vanilla swaps are related to each other through similar convexity adjustment terms.

4.3. Prepayment Options

A major class of instruments that have convexity by design is the broad array of securities associated with mortgages. A mortgage is a Joan secured by the purchaser of a residential or commercial property. Most fixed-rate mortgages have a critical property. They contain the right to prepay the loan. The mortgage receiver has the right to pay the remaining balance of the loan at any time, and incur only a small transaction cost. This is called a prepayment option and introduces negative convexity in mortgage-related securities. In fact, the prepayment option is equivalent to an American style put option written on the mortgage rate R_t. If the mortgage rate R_t falls below a limit R^K, the mortgage receiver will pay back the original amount denoted by N, by refinancing at the new rate R_t. Instead of making a stream of fixed annual interest payments Rtll N, the mortgage receiver has the option (but not the obligation) to pay the annual interest R_{ti} N at same time l_i. The mortgage receiver may exercise this option if $R_{ti} < R_{to}$. The situation is reversed for the mortgage issuer.

The existence of' such prepayment options create negative convexity for mortgage-backed securities (MBS) and other related asset classes. Since the prepayment option involves an exchange of one fixed stream of payments against another fixed stream, it is clear that interest rate swaps playa critical rate in hedging and risk-managing these options dynamically. We will deal with this important topic in Chapter 18.

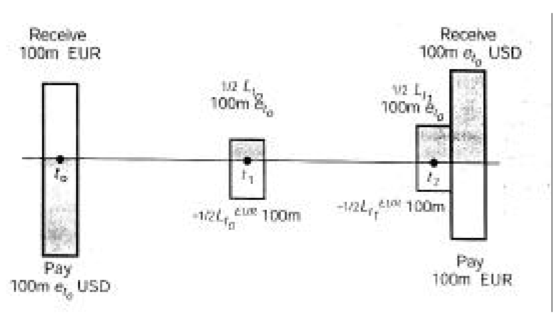
5. A Special Instrument: Quantos

Quanta type financial products farm a major class of instruments where price depends an correlations. At the end of this chapter, we will look at these in detail, and study the financial engineering of quantos by discussing their characteristics and other issues. This can be regarded as another example to the methods introduced in Chapters 8 and 9. We will consider pricing of quantos in Chapter 12.

5.1. A Simple Example

Consider the standard currency swap in Figure 9-7. There are two cash flaws, in two currencies, USD and EUR. The principal amounts are exchanged at the start date and reexchanged at the end date. During the life of the swap, floating payments based an USD Libor are exchanged far floating payments based on EUR Libor. There will be a small known spread involved in these exchanges as well.

The standard currency swap of Figure 9-7 will now be modified in an interesting way. We keep the two floating Libor rates the same, but force all payments to' be made in one currency only, say USD. In other wards, the EUR Libor indexed cash flaws will be paid (received) in USD. This instrument is called a quanta swap, or differential swap. In such a swap, the principal amounts would be in the same currency, and there would be no need to exchange them. Only, net interest rate cash flaws will be exchanged.



Example

Suppose the notional principal is USD30 million. Quotes 011 Libor are as follows:

TENOR	YEN Libor	DOLLAR Libor
3-month	0.055	1.71
6-month	0.185	1.64
12-month	0.065	1.73

In a quanto swap, one party would like to receive 6-month USD Libor and pay 6-month JPY Libor for 1 year. However, all payments are made in USD. For example, if the first settlement is according to the quotes given in the table, in 6 months this party will receive:

$$30{,}000{,}000(.0164)\left(\frac{1}{2}\right) - 30{,}000{,}000(0.00185)\left(\frac{1}{2}\right) - 30{,}000{,}\mathrm{OOO}\left(\frac{1}{2}\right)c \qquad (82)$$

where the c is a constant spread that needs to be determined in the pricing of this quanto swap. Note that the JPY interest rate is applied to a USD denominated principal.

In this type of swap, the two parties are exposed to the risk of interest rate differentials. However, at least one of them is not exposed to currency risk.

Why would anyone be interested in quanto swaps? Note that even after the spread c is included, the interest cost paid in dollars,

$$\text{JPY Libor} + c \qquad (83)$$

may be significantly less than USD Libor rates. This way, the party that receives USD Libor and pays IPY Libor (in USD) may be lowering funding costs substantially. Accordingly, the market would see interest in such quanto swaps when the short ends of the yield curves in two major currencies are significantly different. Banks could then propose these instruments to their clients as a way of "reducing" funding costs. Of course, from the clients' point of view, quanto swaps still involve an interest rate risk and, possibly, an exchange rate risk. If the underlying yield curves shift in unexpected ways, losses may be incurred.

The following example illustrates these from the point of view of British pound and Swiss franc interest rates.

Example

With European economies at a very different point in the trade cycle, corporates are looking to switch their debts into markets offering the cheapest funding. But whereas most would previously have been dissuaded by foreign exchange risk, the emergence of quanto products has allowed them to get the best of both worlds.

With quanto swaps, interest is paid in a different currency to that of the reference index, the exchange rate being fixed at the outset of the swap. As a result, the product can provide exposure to a non-domestic yield curve without the accompanying exchange rate risks.

In recent weeks this type of product has proved increasingly appealing to UK corporates that have entered into a swap in which the paying side is referenced to Swiss Libor but the returns are paid in sterling. Swiss franc Libor is still low relative to sterling Libor and although the corporate ends up paying Swiss Libor plus a spread, funding costs are often still considerably cheaper than normal sterling funding. Deals have also been referenced to German or Japanese Libor..

However, derivatives officials were also keen to point out that quanto products are far from being risk-free. "Given that the holder of the swap ends up paying Swiss Libor plus a spread, the curves do not have to converge much to render the trade uneconomic," said one. (IFR, lssue-l190, July 5 1997.)

5.1.1. Quantos in Equity

The notion of a quanto instrument can be applied in other financial markets. For example, a for-eign investor may want to have exposure to Japanese equity markets without having to incur cur-rency risk. Then, a quanto contract can be designed such that, the gains and losses of an index in Japanese equities are paid annually in the foreign investors' domestic currency, instead of in yen.

5.2. Pricing

The pricing of quanto contracts raises interesting financial engineering issues.[17] We discuss a simple case to illustrate quantos. First, fix the underlying. Assume that we are dealing with a particular foreign currency denominated stock S^*_t. Without loss of generality, suppose the domestic currency is USD, the foreign currency is euro, and the stock is European.

A dollar-based investor would like to buy the stock, and benefit from potential upside in European markets, but, dislikes currency exposure to euro. The investor desires exposure to underlying equity risk only. To accommodate his wish, the bank proposes purchasing the stock via a quanto forward. An expiration date T is chosen, and the current exchange rate EUR/USD e_t is used to calculate the time- T settlement. The forward contract has USD price F_t, and settles according to

$$V_T = (e_t S^*_T - F_t) \quad (84)$$

Here, the VT is the time-T value of the contract. It is measured in the domestic currency, and will be positive, if the stock appreciates sufficiently. Otherwise, it will be negative. The Ft is the forward price of the quanto contract on S^*_T and has to be determined by a proper pricing, strategy

5.3. The Mechanics of Pricing

Suppose the current time is t and a forward quanto contract on S^*_T is written with settlement date

$T = t + \Delta$ Suppose also that at time T there are only three possible states of the world, {? [1], ? [2], ? [3]}. The following table gives the possible values of four instruments, the foreign stock, a foreign deposit, a domestic deposit, and a forward FX contract on the exchange rate e_t.

Time t price	value in ω^1	value in ω^2	value in ω^3
S*t	$S^{*1}_{t+\Delta}$	$S^{*2}_{t+\Delta}$	$S^{*3}_{t+\Delta}$
1USD	$(1+r\Delta)$	$(1+r\Delta)$	$(1+r\Delta)$
$1e_t$	$e^1_{t+\Delta}(1+r^*\Delta)$	$e^2_{t+\Delta}(1+r^*\Delta)$	$e^3_{t+\Delta}(1+r^*\Delta)$
0	$f_t - e^1_{t+\Delta}$	$f_t - e^2_{t+\Delta}$	$f_t - e^3_{t+\Delta}$

In this table, the firsl row gives the value of the foreign stock in the three future states of the world. These are measured in the foreign currency. The second row represents what happens to 1 dollar invested in a domestic savings account. The third row shows what happens when 1 unit of foreign currency is purchased at e_t dollars and invested at the foreign rate r*.

The forward exchange rate f_t is priced as

$$f_t = e_t \frac{1+r\Delta}{1+r^*\Delta} \quad (85)$$

where, e_t is the current exchange rate. In this example, we are assuming that the domestic and foreign interest rates are constant at rand r* respectively. Now consider the quanto forward contract with current price F_t mentioned earlier. The F_t will be determined at time t, and the contract will settle at time $T = t + \Delta$ Depending on which state occurs, the settlement amount will be one of the following:

$$\{(S^{*1}_{t+\Delta} e_t - F_t),\ (S^{*2}_{t+\Delta} e_t - F_t),\ (S^{*3}_{t+\Delta} e_t - F_t)\} \quad (86)$$

These amounts are all in USD. What is the arbitrage-free value of F_t?

We can use three of the four instruments listed to form a portfolio with weights $?_i$, i = 1, 2, 3 that replicate the possible values of $e_t S^*_{t+\Delta}$ at each state exactly. This will be similar to the cases discussed in Chapter 7. For example, using the first three instruments, for each state we can write

$$\lambda_1 s^{*1}_{t+\Delta} e^1_{t+\Delta} + \lambda_2(1+r\Delta) + \lambda_3 e^1_{t+\Delta}(1+r^*\Delta) = S^{*1}_{t+\Delta} e_t \quad (87)$$

$$\lambda_1 s^{*2}_{t+\Delta} e^2_{t+\Delta} + \lambda_2(1+r\Delta) + \lambda_3 e^2_{t+\Delta}(1+r^*\Delta) = S^{*2}_{t+\Delta} e_t \quad (88)$$

$$\lambda_1 s^{*3}_{t+\Delta} e^3_{t+\Delta} + \lambda_2(1+r\Delta) + \lambda_3 e^3_{t+\Delta}(1+r^*\Delta) = S^{*3}_{t+\Delta} e_t \quad (89)$$

In these equations the right-hand side is the future value of the foreign stock measured at current exchange rate. The left-hand side is the value of the replicating portfolio in that state.

These form three equations in three unknowns, and, in general, can be solved for the unknown $?_i$. Once these portfolio weights are known, the current cost of putting the portfolio together leads to the price of the quanto:

$$?_1 S^*_t e_t + ?_2 + ?_3 e_t \quad (90)$$

This USD amount needs to be carried to time T. since the contract settles at T. This gives

$$Ft = [\lambda_1 S^*_t et + \lambda_2 + \lambda_3 et](1 + r\Delta) \quad (91)$$

Example

Suppose we have the following data on tile first three rows of the previous table:

Time t price	value in WI	value in w2	value in W3
100	115	100	90
I USD	(1 + .03Δ)	(1+ .03Δ)	(1 + .03Δ)
I EUR X 0.98	(1+ .03Δ) 1.05	(1 + .03 Δ) 0.98	(1 + .03 Δ) 90

What is the price of tile quanta forward?

We set up the three equations

$$\lambda_1(1.05)115 + \lambda_2(1+0.5\Delta) + \lambda_3 1.05\,(1+.03\Delta) = 0.98\ (115) \quad (92)$$

$$\lambda_1(0.98)100 + \lambda_2(1+0.5\Delta) + \lambda_3\, 0.98\,(1+.03\Delta) = 0.98\ (100) \quad (93)$$

$$\lambda_1(0.90)\,90 + \lambda_2(1+0.5\Delta) + \lambda_3\, 0.90\,(1+.03\Delta) = 0.98\ (90) \quad (94)$$

We select the expiration $\Delta = 1$, for simplicity, and obtain

$$?_1 = 0.78 \quad (95)$$

$$?_2 = 60.67 \quad (96)$$

$$?_3 = -41.53 \quad (97)$$

Borrowing 42 units of foreign currency, lending 61 units of domestic currency, and buying 0.78 units of the foreign stock would replicate the value of the quanto contract at time t + 1. The price of this portfolio at t will be

$$100?_1 O.98 + ?_2 + 0.98?_3 = 96.41 \quad (98)$$

If this is to be paid at time $t + \Delta$, then it will be equal to the arbitrage-free value of F_t:

$$F_t = (1.05)96.41 = 101.23 \quad (99)$$

This example shows that the value of the quanto feature is related to the correlation between the, movements of the exchange rate and the foreign stock. If this correlation is zero, then the quanto wi1l have the same value as a standard forward. If the correlations positive (negative) then the quanto forward will be less (more) valuable than the standard forward. In the example above the exchange rates and foreign stock were positively correlated and the quantoed instrument cost less than the original value of the foreign stock.

5.4. Where Does Convexity Come In?

The discussion of the previous section has shown that, in a simple one period setting with three possible states of the world, we can form a replicating portfolio for the quantoed asset payoffs at a future date. As the number of states increases and time becomes continuous, this type of replicating portfolio needs readjustment. The portfolio adjustments would, in turn, lead to negative or positive trading gains depending on the sign of the correlation, similar to the case of options. This is where volatilities become relevant. In the case of quanto assets there are, at least two risks involved; namely, exchange rate and foreign equity or interest rates. The covariance between these affects pricing as well.

It is due to the trading gains realized during rebalancing, that the quanto feature will have a positive or negative value at time t_0. Thus, quantos form another class of assets where the non negligibility of second order sensitivities leads to dependence of the asset price on variances and covariances.

5.5. Practical Considerations

Ata first glance, quanto assets may appear very attractive to investors and portfolio managers. After all, a contract on foreign assets is purchased and all currency risk is eliminated. Does this, mean we should always quanto?

Here again, some real-life complications are associated with the instrument. First of all, the purchase of a quanto may involve an upfront payment and the quanto characteristics depend on risk premia, bid-ask spreads, and on transaction costs associated with the underlying asset and the underlying foreign currency. These may be high and an approximate hedge using foreign currency forwards may be cheaper in the end.

Second, quanto assets have expiration dates. If, for some unforeseen reason, the contract is unwound before expiration, further costs may be involved. More importantly, if the foreign asset is held beyond the expiration date, the quanto feature would no longer be in effect.

Finally, the quanto contract depends on the correlation between two risk factors, and this correlation may be unstable. Under these conditions, the parties that are long or short the quanto, have exposure to changes in this correlation parame5Yr. This may significantly affect the mark--to-market value of the quanto contracts.

6. Conclusions

Pricing equations depend on one or more risk factors. When the pricing functions are non-linear, replicating portfolios that use linear assets with periodically-adjusted weights, will lead to positive or negative cash flows during the hedging process. If the underlying volatilities and correlations are significant, trading gains from these may exceed the transaction costs implied by periodic rebalancing, and the underlying non-linearity can be traded.

In this chapter we saw two basic examples to this, one from the fixed income sector which made convexity of bonds valuable, and the second from quanto instruments, which also brought in the covariance between risks. The example on quantos is a good illustration of what happens when term structure models depend on more than one factor. In such an environment, the volatilities as well as the covariances between the underlying risks may become important.

References

Two introductory sources discuss the convexity gains one can extract from fixed-income instru-ments. They are Tuckman (2002) and Jegadeesh and Tuckman (1999). The convexity differences between futures and forwards are clearly handled in Hull (2002). The discussion of the quanto feature used here is from Piros (1998), which is in DeRosa (1998). Wilmott (2000) has a nice discussion of quantoed assets as well Hart (1977) is a very good source on this chapter.

OPTION ENGINEERING WITH APPLICATION

Objectives

- On completion of this lesson you will be able to analyze of how to synthetically create payoff diagram for positions that take a view on the direction of markets and on the directions of volatility.

Hello!

This lesson discusses traditional option strategies from the financial engineering perspective and provides market-based examples. The chapter then moves on to discuss exotic options. We are concerned with portfolios and positions that are taken with a precise gain-loss profile in mind. The players consciously take risks in the hope of benefiting or protecting themselves from an expected movement in a certain risk factor. Most investor behavior is of this kind. Investors buy a stock with a higher (systematic) risk, in anticipation of higher returns. A high-yield bond carries a higher default probability, which the bond holder is willing to bear. For all the different instruments, there are one or more risk factors that influence the gains and losses of the position taken. The investor weighs the risks due to potentially adverse movements in these factors against the gains that will result, if these factors behave in the way the investor expected. Some of the hedging activity can be interpreted in a similar way. This lesson deals with techniques and strategies that use options in doing this. We consider classical (vanilla) as well as modern (exotic) options tools.

According to an important theorem in modern finance, if options of all strikes exist, carefully selected option portfolios can replicate any desired gain-loss profile that an investor or a hedger desires. We can synthetically create any asset using a (static) portfolio of carefully selected options. But, financial positions are taken with a payoff in mind. Hence, we start our discussion by looking at payoff diagrams.

1.1. Payoff Diagrams

Let x_t be a random variable representing the time-t value of a risk factor, and let f (x_T) be a function that indicates the payoff of an arbitrary instrument at "maturity" date T, given the value of x_t at time T>t. We call $f(x_T)$ a payoff function. The functional form of f(.) is known if the contract is well defined. It is customary in textbooks to represent the pair $\{f(x_T), x_T\}$ as in Figures 10-1 or 10-1b. Note that, here, we have a nonlinear upward sloping payoff function that depends on the values assumed by x_T only. The payoff diagram in figure 10-1 is drawn in a completely arbitrary fashion, yet, it illustrates some of the general principles of financial exposures. Let us review these.

First of all, for fairly priced exposures that have zero value of initiation, net exposures to a risk factor, x_t, must be negative for some values of the underlying risk. Otherwise, we would be making positive gains, and there would be no risk of losing money. This would be an arbitrage opportunity. Swap-type instruments fall into this category. If, on the other hand, the final payoffs of the contract are non-negative for all values of x_T, the exposure has a positive value at initiation, and to take the position, an upfront payment will have to be made. Option positions have this characteristic.

Second, exposures can be convex, concave, or linear with respect to x_T, and this has relevance for an investor or market professional. The implication of linearity is obvious: the sensitivity of the position to movements in x_T- is constant. The relevance of convexity was discussed in Chapters 8 & 9. With convexity, movements in volatility need to be priced in, and again options are an important category here.

Finally, it is preferable that the payoff functions f (x_T) depend only on the underlying risk x_T, and do not move due to extraneous risks. We saw in Chapters 8 & 9 that volatility positions taken with options may not satisfy this requirement.

1.1.1. Examples of x_t

The discussion thus far dealt with an abstract underlying, x_t. This underlying can be almost any risk the human mind can think of. The following lists some well – known examples of x_t.

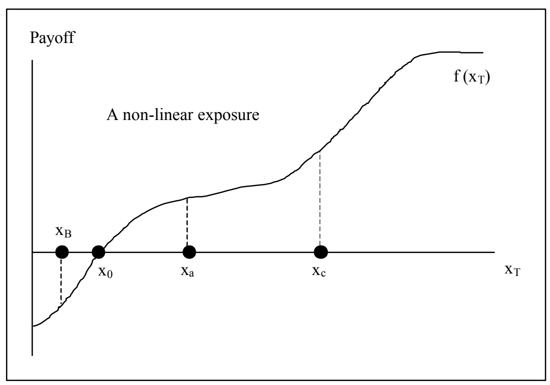


FIGURE 10-1

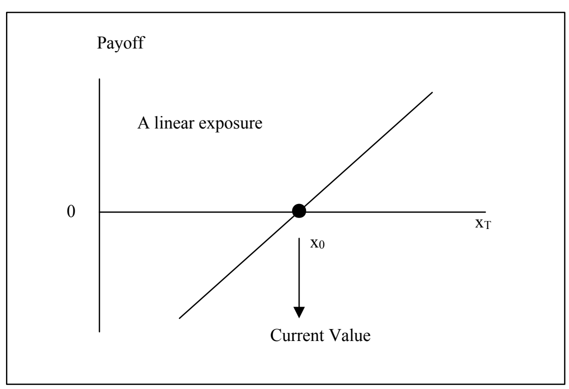


FIGURE 10-1b

- Various interest rates. The best examples are Libor rates and Swap rates. But the commercial paper (CP) rate, the federal funds rate, the index of overnight interest rates (an example of which is EONIA, Euro Over Night Index Average), and many others are also used as reference rates.
- Exchange rates, especially major exchange rates such as dollar-euro, dollar-yen, dollar-sterling ("cable"), and dollar-Swiss frame.
- Equity indices. Here also the examples are numerous. Besides the well-known U.S. indices such as the Dow, Nasdaq, and the S&P500, there are European indices such as CAC40, DAX, and FTSE100, as well as various Asian indices such as the Nikkei 225 and emerging market indices.
- Commodities are also quite amenable to such appositions. Futures on coffee, soybeans, and energy are other examples for x_T.
- Bond price indices. One example is the EMBI + prepared by JPMorgan to track emerging market bonds.

Besides these will-known risks, there are more complicated underlying that, nevertheless are central elements in financial market activity:

1. The underlying to the option positions discussed in this chapter can represent volatility or variance. If we let the percentage volatility of a price, at time t, be denoted by s_t, then the time T value of the underlying x_T may be defined as

$$X_T = \int_t^T s^2_u S^2_u du \qquad (1)$$

   Where S_t may be any risk factor. In this case, x_T represents the total variance of S_t during the interval [t, T]. Volatility is the square root of x_T.
2. The correlation between two risk factors can be traded in a similar way.
3. The underlying, x_t, can also represent the default probability associated with a counterparty or instrument. This arises in the case of credit instruments.
4. The underlying can represent the probability of an extraordinary event happening. This would create a "Cat" instrument that can be used to buy insurance against various catastrophic events.
5. The underlying x_t, can also be a non-storable item such as electricity, weather, or bandwidth.

Readers who are interested in the details of such contracts, or markets should consult Hull (2002). In this chapter, we limit our attention to the engineering aspects of option contracts.

1. Option Strategies

We divide the engineering of option strategies into two broad categories. First, we consider the classical option-related methods. These will cover strategies used by market makers, as well as retail investors. They will themselves be divided into two groups, those that can be labeled directional strategies, and those that relate to views on the volatility of some underlying instrument. The second category involves the exotic options, which we consider as more efficient and, sometimes cheaper alternatives to the classical option strategies. The underlying risks can be any of those mentioned in the previous section.

1.1. Synthetic long and Short Positions

We begin with strategies that utilize options essentially as directional instruments, starting with the creation of long and short positions on an asset. Options can be used to create these positions synthetically.

Consider two plain vanilla options written on a forward price F_t of a certain asset. The first is a short put, and the second a long call, with prices P(t) and c(t) respectively, as shown in Figure 10-2. the options have th same strike price K, and the same expiration time T. Assume that the Black – Scholes conditions hold, and that both options are of European style. Importantly, the underlying asset does not have any payouts during [t, T]. Also, suppose the appropriate short rate to discount future cash flows is constant at r.

Now consider the portfolio

$$\{1 \text{ Long K-Call, } 1 \text{ Short K-Put}\} \qquad (2)$$

At expiration, the payoff from this portfolio will be the vertical sum of the graphs in figure 10-2 and is as shown in figure 10-3. This looks like the payoff function of a long forward contract entered into at K. if the options were at-the-money (ATM) at time t, the portfolio would exactly duplicate the long forward position and hence would be an exact synthetic. But, there is a close connection between this portfolio and the forward contract, even when the options are not ATM.

At expiration time T, the value of the portfolio is

$$C(T) - P(T) = F_T - K \qquad (3)$$

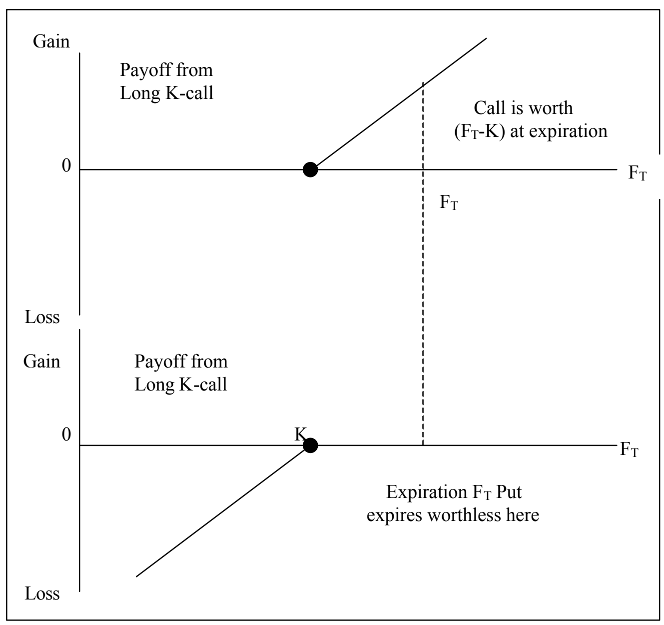


FIGURE 10-2

Where F_T is the time – T value of the forward price. This equation is valid because at T, only one of the two options can be in-the-money. Either the call option has a value of $F_T - K$ while the other is worthless, or the put is in-the-money and the call is worthless, as shown in Figure 10-2. Subtract the time-t forward price, F_t, from both sides of this equation to obtain

$$C(T) - P(T) + (K - F_t) = F_T - F_t \quad (4)$$

This expression says that the sum of the payoffs of the long call and the short put plus $(K - F_t)$ units of cash should equal the time T gain or loss on a forward contract entered into at F_t, at time t.

But the forward contract has zero value at t. Thus, the time t value of the portfolio,

$$\{1 \text{ Long k-Call, 1 Short K-Put, } e^{-r(T-t)}(K - F_t) \text{ Dollars}\} \quad (5)$$

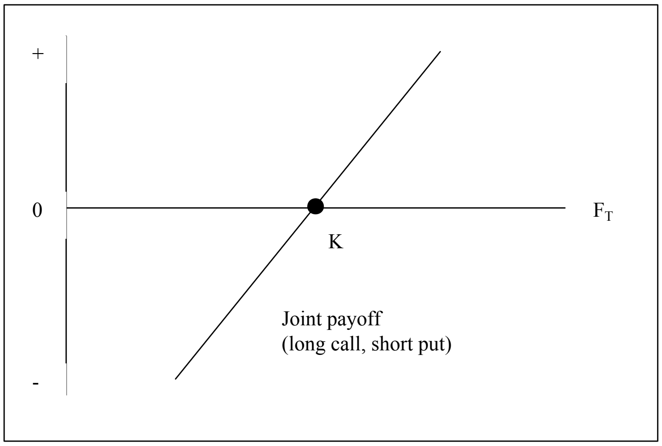


FIGURE 10-3

Should be zero at t, since credit risks and the cash flows generated by the forward and the replicating portfolio are the same. This implies that

$$D(t) - P(t) = e^{-r(T-t)}(F_T - K) \quad (6)$$

This relationship is called put-call parity. It holds for European options. It can be expressed in terms of the spot price, S_t, as well. Assuming zero storage costs, and no convenience yield.

$$F_t = e^{r(T-t)} S_t \quad (7)$$

Substituting in the preceding equation gives:

$$C(t) - P(t) = (S_t - e^{-r(T-t)} K) \quad (8)$$

Put-call parity can thus be regarded as another result of the application of contractual equations, where options and cash are used to create a synthetic for the S_t. This situation is shown in Figure 10-4.

2.1.1. An Application

Option market makers routinely use the put-call parity in exploiting windows of arbitrage opportunities. Using options, market makers construct synthetic futures positions, and then trade them against futures contracts. This way, small and temporary differences between the synthetic and the true contract are converted into "risk less" profits. In this section we discuss an example.

Suppose, without any loss of generality, that a stock is trading at

$$S_t = 100 \quad (9)$$

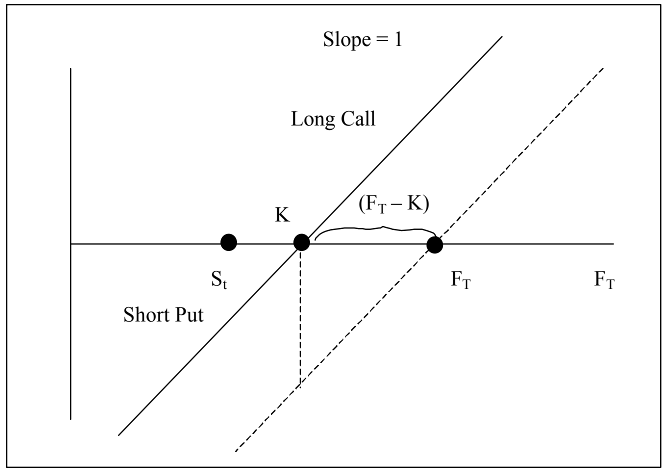


FIGURE 10-4

And that, the market maker can buy and sell at-the-money options that expire in 30 days. Suppose also that, the market maker faces a funding cost 5%. The stock never pays dividends, and there are no corporate actions.

Also, and this is the real-life part, the market maker faces a transaction cost of 20 cents per option and a transaction cost of 5 cents per stock. Finally, the market maker has calculated that to be able to continue operating, he or she needs a margin of .25 cents per position. Then, we can apply put-call parity and follow the convention strategy displayed in Figure 10-5.

Borrow necessary funds overnight for 30 days, and buy the stock at price S_t. At the same time, sell the S_t-call and buy the S_t-put that expire in 30 days, to obtain the position shown in figure 10-5.

The position is fully hedged, as any potential gains due to movement in S_t, will cover the potential losses. This means that the only factors that matter are the transaction costs and any price differentials that may exist between the call and the put. The market maker will monitor the difference between the put and call premiums and take the arbitrage position shown in Figure 10-5, if this difference is bigger than the total cost of the conversion.

Example

Suppose S_t = 100, and 90 – day call and put options trade actively. The interest cost is 5%. A market maker has determined that the call premium. C(t), exceeds the put premium, P(t), by \$2.10:

$$C(t) - P(t) = 2.10 \quad (10)$$

The stock will be purchased using borrowed funds for 90 days, and the ATM put is purchased and held until expiration, while the ATM call is sold. This implies a funding cost of

$$100(.05)(90/360) = \$1.25 \quad (11)$$

Add all the costs of the conversion strategy:

Cost per security	$
Funding Cost	1.25

Stock purchase	0.05
Put purchase	0.20
Call sale	0.20
Operating costs	0.25
Total cost	1.95

The market maker incurs a total cost of \$1.95. It turns out that under these conditions, the net cash position will be positive:

$$\text{Net profit} = 2.10 - 1.95 \tag{12}$$

And the position is worth taking.

If, in the example just discussed, the put-call premium difference is negative, then the market maker can take the opposite position, which would be called a reversal?

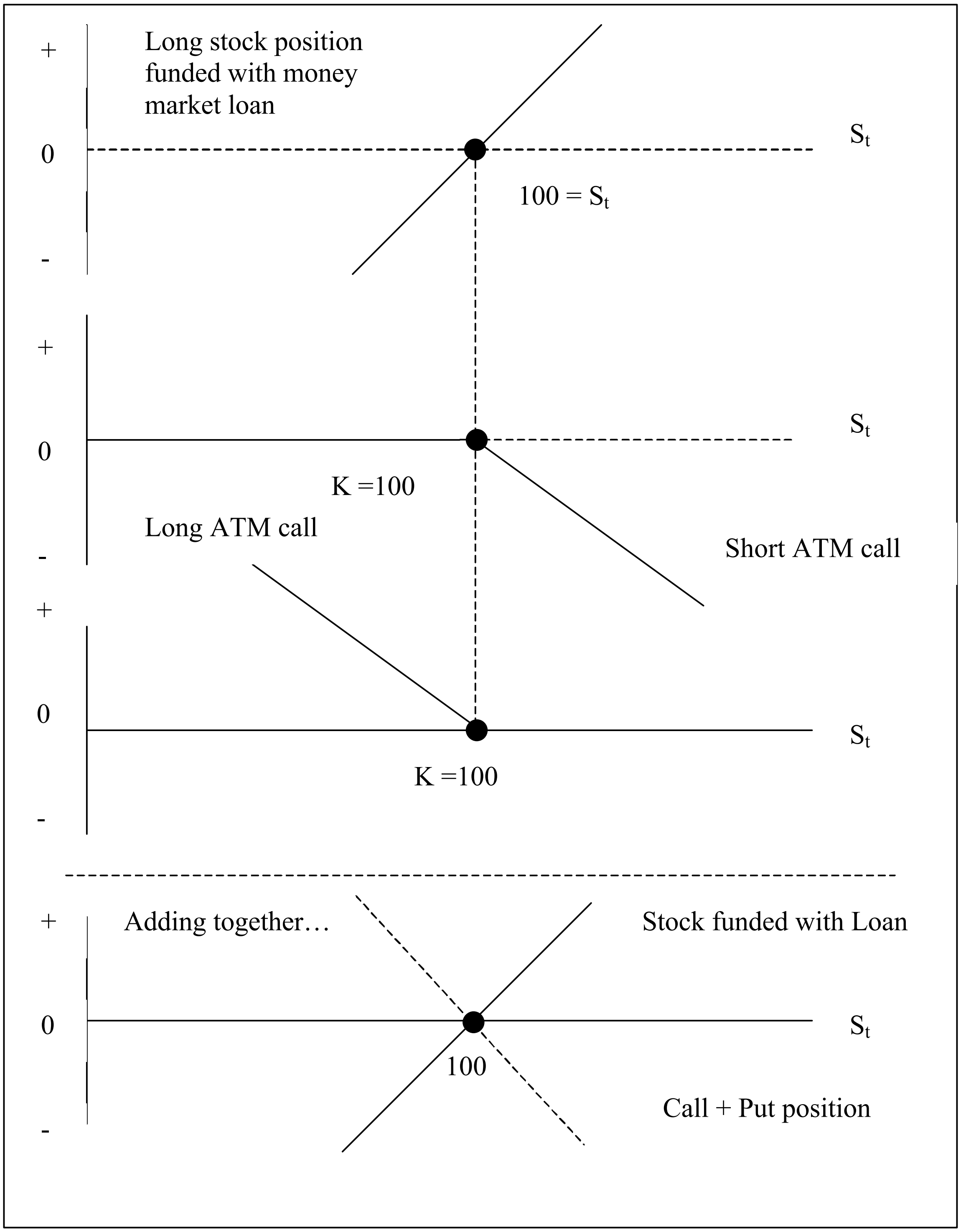


FIGURE 10.5

2.1.2. Arbitrage Opportunity?

An outside observer may be surprised to hear that such "arbitrage" opportunities exist, and that, they are closely monitored by market makers on the trading floor. Yet, such opportunities are available only to market makers on the floor, and may not even constitute arbitrage in the usual sense.

This is because of the following. (1) Off-floor investors pay much higher transactions cost than the on-floor market makers. Then, total costs of taking the position are too high, and may prohibit such positions for off-floor investors. (2) Off-floor investors cannot really make a simultaneous decision to buy (sell) the underlying, and buy or sell the implied puts or calls to construct the strategy. By the time these strategies are communicated to the floor, prices could move. (3) Even if such opportunities are found, net gains are often too small to make it worthwhile to take such positions sporadically. It is, however, worthwhile to a market maker who specializes in these activities. (4) Finally, there is also a serious risk associated with these positions, known as the pin risk.

2.1. A Remark on the Pin Risk

It is worthwhile to discuss pin risk in more detail, since similar risks arise in hedging and trading some exotic options as well. Suppose we put together a conversion at 100, and waited 90 days until expiration, to unwind the position. The positions will expire some 90 days later during a Friday. Suppose at expiration S_t is exactly 100. This means that the stock closes exactly at the strike price. This leads to a dilemma for the market maker.

The market maker owns a stock. If he or she does not exercise the long put, and, if the short call is not assigned (i.e., if he or she does not get to sell at A exactly), then the market maker will have an open long position in the stock during the weekend. These risks may not be great for an end investor who takes such positions occasionally. But they may be substantial for a professional trader who takes such positions occasionally. But they may be substantial for a professional trader who depends on these positions, and there is no easy way out of this dilemma. This type of risk is known as the pin risk.

The main cause of the pin risk is the kink, at $S_T = K$, in the expiration payoff. A kink indicates a sudden change in the slope – for a long call, from zero to one or vice versa. This means that, even with small movements in S_t, the hedge ratio can be either zero or one, and the market maker may be caught significantly off guard. If the slope of the payoff diagram changed smoothly, them the required hedge would also change smoothly. Thus, a risk similar to pin risk may arise whenever the delta of the instrument shows discrete jumps.

2.2. Risk Reversals

A more advanced version of the synthetic long and short futures positions is known as risk reversals. These are liquid synthetics especially in the foreign exchange markets, where they are traded as a commodity. Risk reversals are directional positions, but, differ in more than one way from synthetic long-short futures positions discussed in the previous section.

The idea is again to buy and sell calls and puts in order to replicate long and short futures positions – but this time using options with different strike prices. Figure 10-6 shows an example. The underlying is S_t. The strategy involves a short put struck at K_1, and a long call with strike K_2. Both options are out-of-the-money initially, and the S_t satisfies

$$K_1 < S_t < K_2 \tag{13}$$

Since strikes can be chosen such that the put and call have the same premium, the risk reversal can be constructed so as to have zero initial price.

By adding vertically the option payoffs in the top portion of Figure 10-6, we obtain the expiration payoff shown at the bottom of the figure. If, at expiration, S_T is between K_1 and K_2, the strategy has zero payoff. If, at expiration, $S_T < K_1$, the risk reversal loses money, but under $K_2 < S_T$, it makes money. Clearly, what we have here is similar to a long position but the position is neutral for small movements in the underlying starting from S_t. If taken naked, such a position would imply a bullish view on S_t.

We consider an example from foreign exchange (FX) markets where risk reversals are traded as commodities.

Example

Twenty-five delta one-month risk reversals showed a stronger bias in favor of euro calls (dollar puts) in the last two weeks after the euro started to strengthen against the greenback.

Traders said market makers in EUR calls were buying risk reversals expecting further euro upside. The one-month risk reversal jumped to 0.91 in favor of euro calls Wednesday from 0.3 three weeks ago. Implied volatility spiked across the board. One-month volatility was 13.1% Wednesday from 11.78% three weeks ago as the euro appreciated to USD1.0215 from USD1.0181 in the spot market.

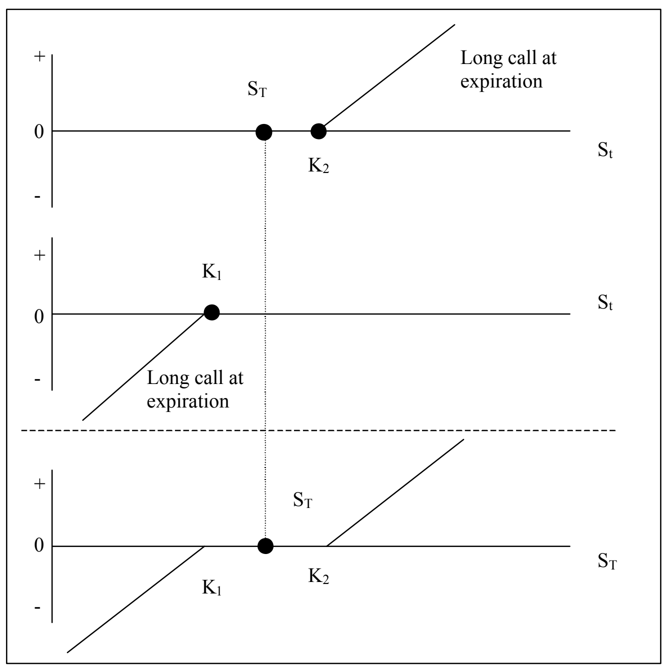


FIGURE 10-6

The 25-delta risk reversals mentioned in this reading are shown in Figure 10-7a. The risk reversal is constructed using two options, a call and a put. Both options are out-of-the-money and have a "current" delta of 0.25. According to the reading, the 25-delta EUR call is more expensive than the 25-delta EUR put.

2.1.1. Uses of Risks Reversals

Risk reversals can be used as "cheap" hedging instruments. Here is an example.

Example

A travel company in Paris last week entered a zero-cost risk reversal to hedge U.S. Dollar exposure to the USD. The company needs to buy dollars to pay suppliers in the U.S., China, Indonesia, and South America.

The head of treasury said it bought dollar calls and sold dollar puts in the transaction to hedge 30% of its USD200 – 300 million dollar exposure versus the USD. The American-style options can be exercised between November and May and have a notional of USD10 -20 million.

The company entered a risk reversal rather than buying a dollar call outright because it was cheaper. The head of treasury said the rest of its exposure is hedged using different strategies, such as buying options outright.

Have we had a corporation that has EUR receivables from tourists going abroad, but, needs to make payments to foreigners in dollars? Euros are received at time t, and dollars will be paid at some future date T, with t<T. The risk reversal is put together as a zero cost structure, which means that the premium collected from selling the put (on the USD) is equal to the call premium on the USD. For small movements in the exchange rate, the position

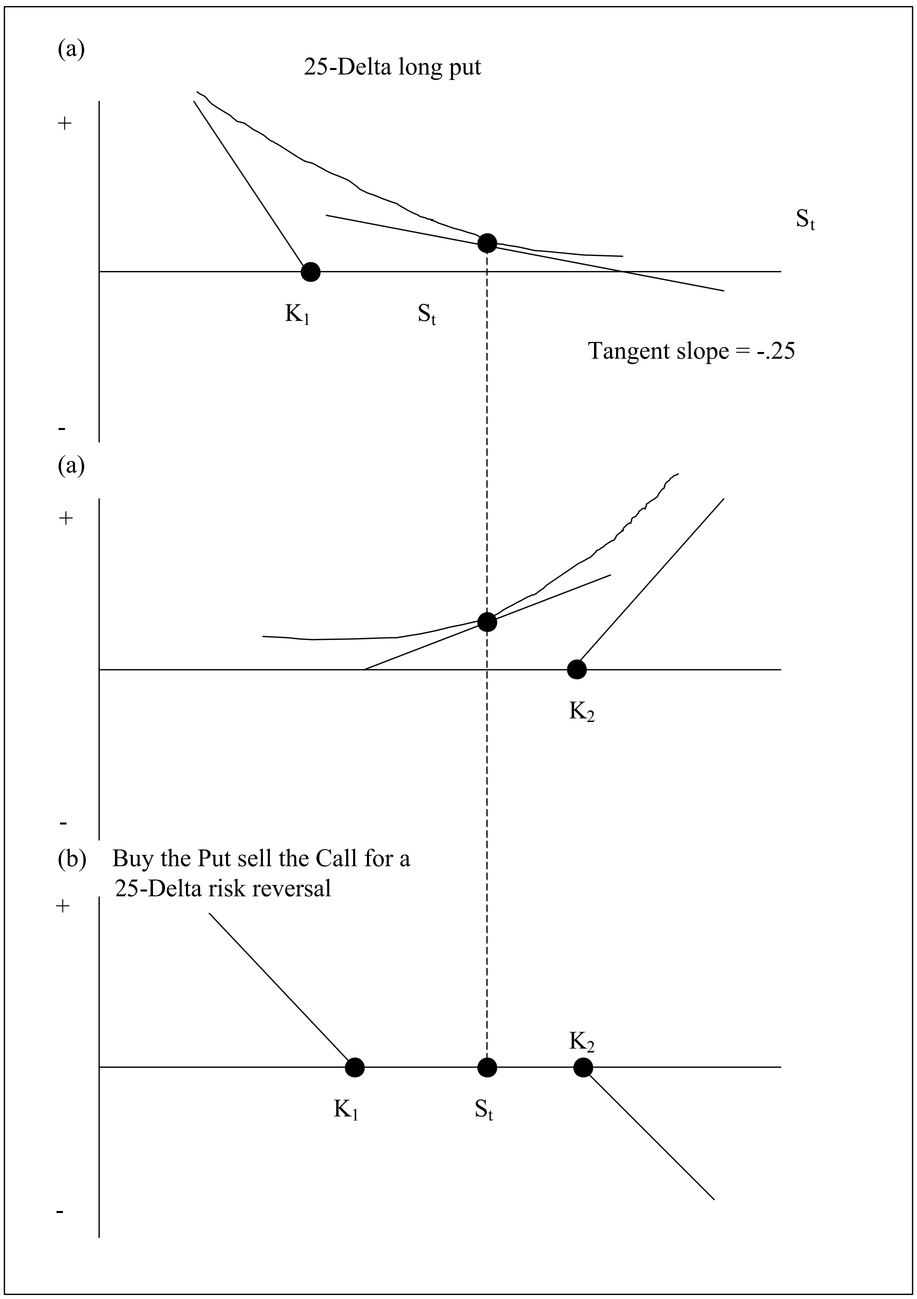


FIGURE 10-7

is neutral, but for large movements it represents a hedge similar to a futures contract.

Of course, such a position could also be taken in the futures market. But one important advantage of the risk reversal is that it is "composed" of options, and hence involves, in general, no daily mark-to-market adjustments.

2.4. Yield Enhancement Strategies

The class of option strategies that we just studied thus far is intended for creating synthetic short and long futures positions. In this section, we consider option synthetics that are said to lead to yield enhancement for investment portfolios.

2.4.1. Call Overwriting

The simplest case is the following. At time t, an investor takes a long position in a stock with current price S_t, as shown in Figure 10-8. If the stock price increases, the investor gains; if the price declines, he or she loses. The investor has, however, a subjective expected return, $\hat{R}_t$, for an interval of time Δ, that can be expressed as

$$\hat{R}_t = E_t^{\hat{P}}\left[\frac{S_t + \Delta - S_t}{S_t}\right] \qquad (14)$$

Where $\hat{p}$ is a subjective conditional probability distribution for the random variable $S_t + \Delta$. According to the formula, the investor is expecting a gain $\hat{R}_t$ during period Δ. The question is whether we can provide a yield-enhancing alternative to this investor. The answer depends on what we mean by "yield enhancement."

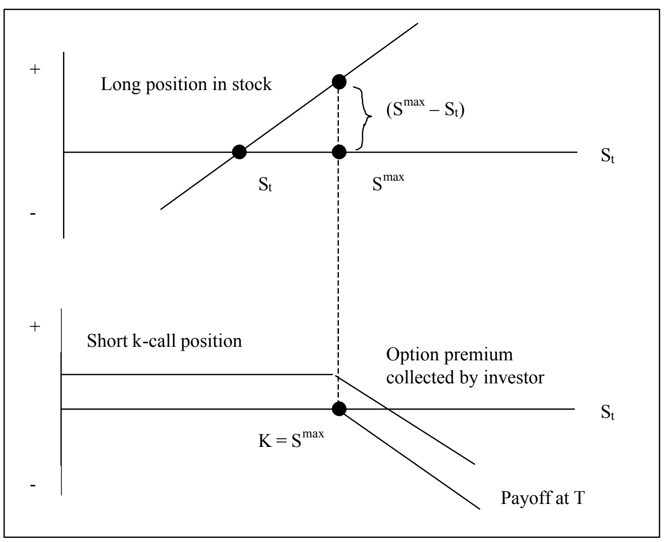


FIGURE 10-8

Suppose we ask the investor the following question: "What is the maximum gain that you would like to make on this tock position?" Assume for the time being that the investor indicates that, S_{max} is the price he or she will be most happy to sell the stock at, and realize the "maximum" desired gain:

$$(S^{max} - S_t) \qquad (15)$$

Next, consider a call option $C(t)^{max}$ that has the strike

$$K = S^{max} \qquad (16)$$

And that expires at $T = t + \Delta$. This options sells for $C(t)^{max}$ at time t. We can then recommend the following portfolio to this investor:

$$\text{Yield enhanced portfolio} = \{\text{Long } S_t, \text{ Short } C(t)^{max}\} \qquad (17)$$

Assuming zero interest rates, at time $T = t + \Delta$, this portfolio has the following value, $V_{t+\Delta}$:

$$V_{t+\Delta} = \begin{cases} C(t)^{max} + S_{t+\Delta} & \text{Option not exercised} \\ C(t)^{max} + S_{t+\Delta} - (S_{t+\Delta} - (S_{t+\Delta} - S^{max}) = C(t)^{max} & \text{Option exercised} \end{cases} \qquad (18)$$

According to this, if at expiration, the price stays below the level S^{max}, the investor "makes" an extra $C(t)^{max}$ dollars. If $S_{t+\Delta}$ exceeds the S^{max}, the option will be exercised, the gains will be truncated at $S^{max} + C(t)^{max}$. But, this amount is higher than the price at which this investor was willing to sell the stock according to his or her subjective preferences. As a result, the option position has enhanced the "yield" of the original investment. However, it is important to realize that what are being enhanced is not the objective risk-return characteristics, but instead, the subjective expected returns of the investor.

Figure 10-8 shows the situation graphically. The top portion is the long position in the stock. The bottom profile is the payoff of the short call, written at strike S^{max}. If $S_{t+\Delta}$ exceeds this strike, the option will be in-the-money and the investor will have to surrender his or her stock, worth $S_{t+\Delta}$ dollars, at a price of S^{max} dollars. But, the investor was willing to sell at S^{max} anyway. The sum of the two position is illustrated in the final payoff diagram in Figure 10-9.

The strategy is called call overwriting and is frequently used by some investors. The following reading illustrates one example. Fund managers who face a stagnant market use call over-writing to enhance yields.

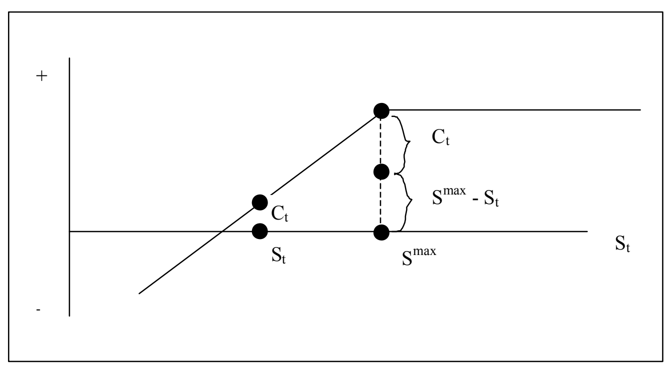


FIGURE 10-9

Example

Find manager motivation for putting on options strategies ahead of the Russell indices annual rebalance next month is shifting, say some options strategists.

"The market has had no direction since May last year," said a head of equity derivatives strategy in New York. Small cap stocks have only moved up slightly during the year, he added.

Fund managers are proving increasingly willing to test call overwriting strategies for the rebalance as they seek absolute returns, with greater competition from hedge funds pushing derivatives strategy said. Employing call overwriting strategies – even though they suppress volatility on the downside. As such, it can help managers enhance their returns. (IFR, Issue 1433, May 11, 2002.)

The situation described in this reading is slightly more complicated and would not lend itself to the simple call overwriting position discussed earlier. The reading illustrates the periodic and routine rebalancing that needs to be performed by fund managers. Many funds "track" well-known indices. But, these indices are periodically revised. New names enters, others leave, at known dates. A fund manager, who is trying to track a particular index, has to rebalance his or her portfolio as indices are revised.

2. Volatility – Based Strategies

The first set of strategies dealt with directional uses of options. Option portfolios combined with the underlying, were used to take a view on the direction of the underlying risk. Now we start looking at the use of options from the point of view of volatility positioning. The strategy used in putting together volatility positions in this section is the following. First, we develop a static position that eliminates exposure to market direction. This can be done using straddles and their cheaper version, strangles. Second, we combine strangle and straddle portfolios to get more complicated volatility positions, and to reduce costs.

Thus, the basic building blocks of volatility positions considered in this chapter are straddles and strangles. The following example indicates how an option position is used to take a view on volatility, rather than the price of the underlying.

Example

An Italian bank recommended the following position to a client. We will analyze what this means for the client's expectations [views] on the markets. First we read the episode.

"A bank last week sold 4% out-of-the-money puts and calls on ABC stock, to generate a premium on behalf of an institutional investor. The strangle had a tenor of six weeks... The strategy generated 2.5% of the equity's spot level in premium.

At the time of the trade, the stock traded at roughly USD1, 874.6. Volatilities were at 22% when the options were sold. ABC was the underlying, because the investor does not believe the stock will move much over the coming weeks and thus is unlikely to break the range and trigger the options". (Based on an article in Derivatives Week)

Figure 10-10 shows the payoff diagram of these options position at expiration. Adding the premiums received at the initial point we get the second diagram in the bottom part of the figure. This should not be confused with the anticipated payoff of the client. Note that the eventual objective of the client is to benefit from volatility realizations. The option position is only a vehicle for doing this.

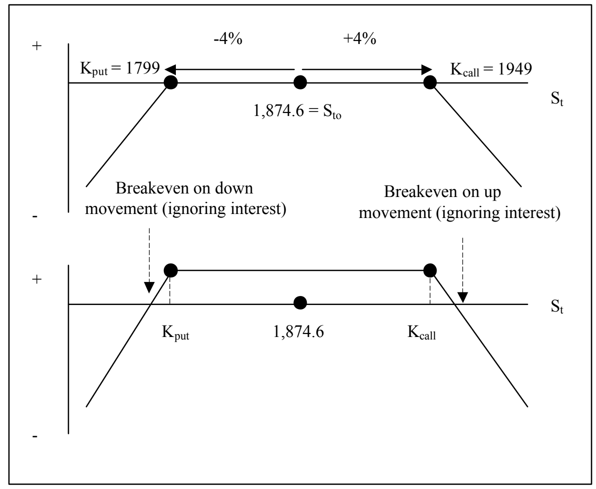


FIGURE 10-10

We can discuss this in more detail. The second part of Figure 10-10 shows that at expiration, the down and up breakeven points for the position are 1762 and 1987, respectively. These are obtained by subtracting and adding the \$37.5 received from the strangle position, to the respective strike prices.

But, the reading also gives the implied volatility in the market. From here we can use the square root formula and calculate the implied volatility for the period under consideration

$$\sigma S_t \Delta = 0.22 \sqrt{\frac{6}{52}}\ 1874.6 = 140.09 \qquad (19)$$

Note that the breakeven points are set according to 4% movements toward either side, whereas the square root formula gives 7.5% expected movements to either side.

According to this, the client who takes this position expects the realized volatility to be significantly less than the 7.5% quoted by the market. In fact, the client expects volatility to be somewhat less than 4%.

This brings us to a formal discussion of strangles and straddles, which form the main building blocks for classical volatility positions.

3.1 Strangles

Assume that we sell (buy) two plain vanilla. European-style options with different strikes on the asset S_{-t}. The first is a put, and has strike K_p; the second is a call, and has strike K_c, with $K_p < K_c$. Suppose at the time of purchase, $K_p < S_{to} < K_c$. The expiration date is T. This position is known as a strangle, and an example for its use in the market was shown earlier. Because these options are sold, the seller collects a premium, at time t, o f

$$C(t) + P(t) \qquad (20)$$

The position makes money if, by expiration, S_t has moved by a "moderate" amount, otherwise the position loses money. Clearly, this way of looking at a strangle suggests that the position is static. A typical short strangles expiration payoff is shown in figure 10-11. The same figure indicates the value of the position at time t, when it was initially put in place.

3.1.1. Uses of Strangles

We give an example for the use of strangles. The example is from foreign exchange markets. Note the switch in terminology. Instead of talking about options that are out-of-the-money by k% of the strike, the episode uses the terminology "10-delta options." This is the case because, as mentioned earlier, FX markets like to trade 10-delta, 25-delta options, and these will be more liquid than, say, an arbitrarily selected k% out-of-the-money option.

Examples

A bank is recommending its clients to sell one-month 10-delta euro/dollar strangles to take advantage of low holiday volatility. The strategists said the investors should sell one-month strangles with puts struck at USD.8380 and calls struck at USD0.9350. This will generate a premium of 0.3875% of the notional size. Spot was trading at USD0.8840 when the trade was designed last week. Euro/dollar sport was at USD0.8786 on Wednesday.

The bank thinks this is a good time to put the trade on because implied volatility traditionally falls over Christmas and New Years, which means spot is likely to stay in this range. (Based on Derivatives Week)

This is a straightforward use of strangles, and is shown in Figure 10-12. According tot the strategist, the premium associated with the FX options implies a volatility that is higher than the expected future realized volatility during the holiday season, due to seasonal factors. If so, the euro/dollar exchange rate is likely to be range-bound, and the options used to create the strangles will expire unexercised.

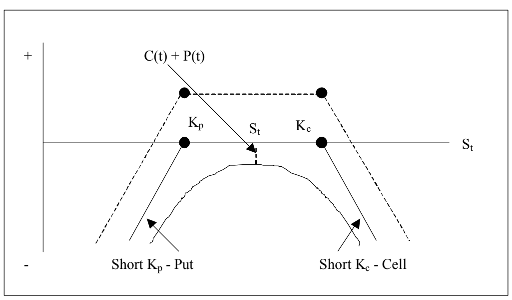


FIGURE 10-11

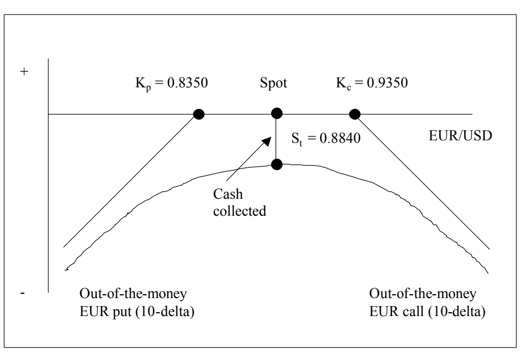


FIGURE 10-12

3.2. Straddle

A straddle is similar to a strangle, except that the strike prices, K_p and K_c, of the constituent call and put options old (bought), are identical:

$$K_p = K_c \tag{21}$$

Let the underlying asset be S_t, and the expiration time be T. The expiration payoff and time value of a long straddle are shown in Figure 10-13. The basic configuration is similar to a long strangle. One difference is that, a straddle will cost more. At the time of purchase, an ATM straddle is more convex that an ATM strangles, and hence has "maximum" gamma.

3.2.1. Static or Dynamic Position?

It is worthwhile to emphasize that the strangle or straddle positions discussed here are static, in the sense that, once the positions are taken, they are not delta-hedged. However, it is possible to convert them into dynamic strategies. To do this, we would delta-hedge the position dynamically. At initiation, an ATM straddle is automatically market-neutral, and the associated delta is zero. When the price moves up, or down, the delta becomes positive, or negative. Thus, to maintain a market-neutral position, the hedge needs to be adjusted periodically.

Note a major difference between the static and dynamic approaches. Suppose we take a static straddle position, and S_t fluctuates by small amounts very frequently and never leaves the region $[S_1, S_2]$ shown in Figure 10-14. Then, the static position will lose money, while the dynamic delta-hedged position may make money, depending on the size and frequency of oscillations in S_t.

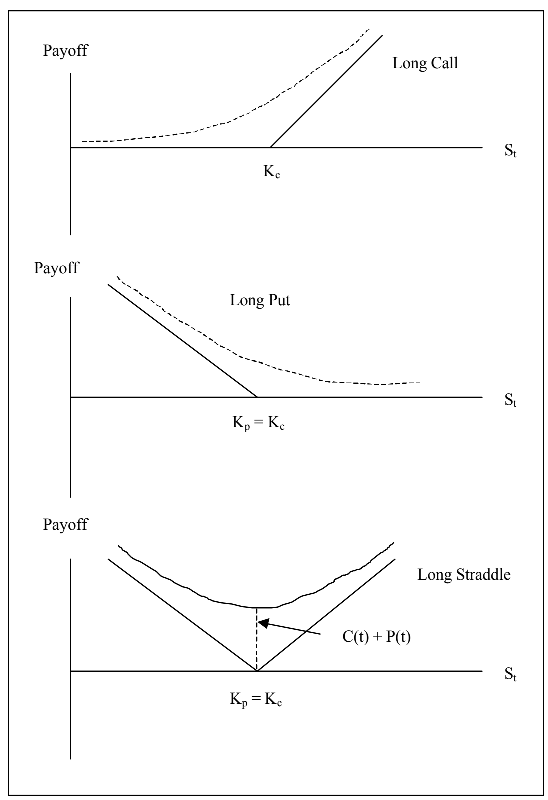


FIGURE 10-13

3.2. Butterfly

A butterfly is a position that is built using combinations of strangles and straddles. Following the same idea used throughout the book, once we develop strangle and straddle payoffs as building blocks, we can then combine them to generate further synthetic payoffs. A long butterfly position is shown in Figure 10.15. The figure implies the following contractual equation:

Long butterfly = Long ATM straddle + Short k% out-of-the-money strangle (22)

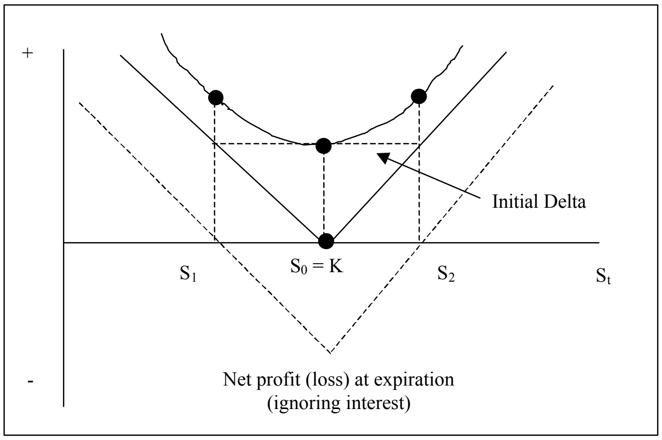


FIGURE 10-14

This equation immediately suggests one objective behind butterflies. By selling the strangles, the trader is, in fact, lowering the cost of buying the straddle. In the case of the short butterfly, the situation is reversed:

Short butterfly = Short ATM straddle + Long k% out-of-the-money strangle (23)

A short straddle generates premiums, but, has unlimited downside. This may not be acceptable to a risk manager. Hence, the trader buys a strangle to limit these potential losses. But this type of insurance involves costs and the net cash receipts become smaller. The following shows a practical use of the short butterfly strategy.

Example

As the Australian dollar continues to strengthen on the back of surging commodity prices, dealers are looking to take advantage of an anticipated lull in the currency's bull run by putting in place butterfly structures. One structure is a three-month butterfly trade. The dealer sells an at-the-money Aussie dollar call and put against the U.S. dollar, while buying an Aussie call struck at AUD0.682 and buying puts struck at AUD0.6375. The structure can be put in place for a premium of 0.3% of notional, noted one trader, adding that there is value in both the puts and the calls. (Based on an article in Derivatives Week)

This structure can also be put in place by making sure that the exposure is vega-neutral.

4. Exotics

Up to this point, the chapter has dealt with option strategies that used only plain vanilla calls and puts. The more complicated volatility building blocks, namely straddles and strangles, were generated by putting together plain vanilla options with different strike prices or expiration.

But the use of plain vanilla options to take a view on the direction of markets or to trade volatility may be considered by some as "outdated". There are now more practical ways of accomplishing similar objectives.

The general principle is this. Instead of combining plain vanilla options to create desired payoff diagrams, lower costs, and reach other objectives, a trader would directly design new option contracts that can do similar things in a "better" fashion. Of course, these new contracts imply a hedge that is, in general, made of the underlying plain vanilla options, but the new instruments themselves would sell as exotic options. Before closing this chapter, we would like to introduce further option strategies that use exotic options as building blocks. We will look at

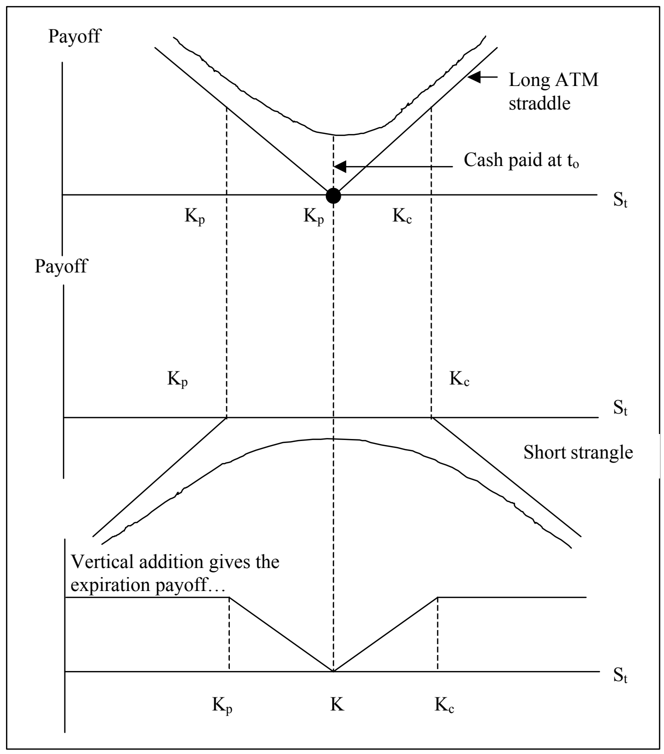


FIGURE 10-15

a limited number of exotics, although there are many others that we relegate to the exercises at the end of the chapter.

4.1. Binary, or One-Touch Options

To understand binary options, first remember the static strangle and straddle strategies. The idea was to take a long (short) volatility position, and benefit if the underlying moved more (less) than what the implied volatility suggested. Binary options form essential building blocks for similar volatility strategies, which can be implemented in a cheaper and perhaps more efficient way. Also, binary options are excellent examples of option engineering. We begin with a brief description of a European style binary option.

4.1.1. A Binary Call

Consider a European call option with strike K and expiration time T. S_t denotes the underlying risk. This is a standard call, except that, if the option expires at or in-the-money, the payoff will be either (1) a constant cash amount or (2) a particular asset. In this section, we consider binaries with cash payoffs only.

Figure 10-16 shows the payoff structure of this call whose time-t price is denoted by $C^{bin}(t)$. The time T payoff can be written as

$$C^{bin}(T) = \begin{cases} R & \text{If } K \leq S_T \\ 0 & \text{Otherwise} \end{cases} \qquad (24)$$

According to this, the binary call holder receives the cash payment R as long as S_T is not less than K at time T. Thus, the payoff has a R-or-nothing binary structure. Binary puts are defined in a similar way.

The diagram in Figure 10-16 shows the intrinsic value of the binary. What would the time value of the binary option look like? It is, in fact, easy to obtain a closed-form formula that will price binary options. Yet, we prefer to answer this question using financial engineering. More precisely, we first create a synthetic for the binary option. The value of the synthetic should then equal the value of the binary.

The logic in forming the synthetic is the same as before. We have to duplicate the final payoffs of the binary using other (possibly liquid) instruments), and make sure that the implied cash flows and the underlying credit risks are the same.

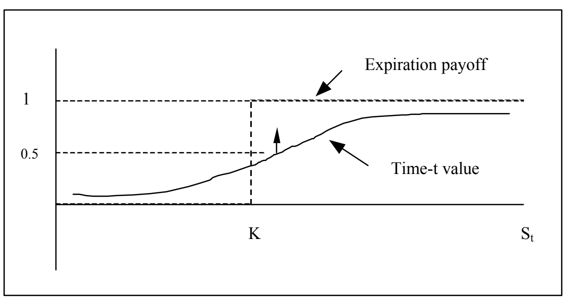


FIGURE 10-16

4.1.2. Replicating the Binary Call

Final payoffs of the binary are displayed by the step function shown in figure 10-16. Now, make two additional assumptions. First, assume that the underlying S_t is the price of a futures contract traded at an exchange, and that the exchange has imposed a minimum tick rule such that, given S_t, the next instant's price, $S_{t+\Delta}$, can only equal

$$S_{t+\Delta} = S_t \pm ih \qquad (25)$$

Where i is asset integer, and h is the minimum tick chosen by the exchange. The parameter Δ represents a small interval. Second, we assume without any loss of generality that

$$R = 1 \qquad (26)$$

Under these conditions, the payoff of the binary is a step function that shows a jump of size 1 at $S_T = K$.

It is fairly easy to find a replicating portfolio for the binary option under these conditions. Suppose the market maker buys one vanilla European call with strike K, and, at the same time, sells one vanilla European call with strike K + h on the S_t. Figure 10-17 shows the time T payoff of this portfolio. The payoff is similar to the step function in Figure 10-16, except that the height is h, and not 1. But this is easy to fix. Instead of buying and selling 1 unit of each call, the market maker can buy and sell 1/h units. This implies the approximate contractual equation.

Binary call, Strike K	= ~	Long 1/h units Of vanilla K-call	+	Short 1/h units of vanilla (K + h)-call

The existence of a minimum tick makes this approximation a true equality, since $|S_t - S_{t+\Delta} < h|$ cannot occur due to minimum tick requirements. We can use this contractual equation and get two interesting results.

4.1.3. Delta and Price of Binaries

There is an interesting analogy between binary options and the delta of the constituent plain vanilla counterparts. Let the price of the vanilla K and K +h calls be denoted by C k (t) and Ck+h (t), respectively. Then, assuming that the volatility parameter ó does not depend on K, we can let h à 0 in the previous contractual equation, and obtain the exact price of the binary $C^{bin}(t)$, as

$$C^{bin}(t) = \lim_{h \to 0} \frac{C^k(t) - C^{k+h}(t)}{h} \qquad (27)$$

$$= \frac{\partial C^k(t)}{\partial K} \qquad (28)$$

assuming that the limit exists.

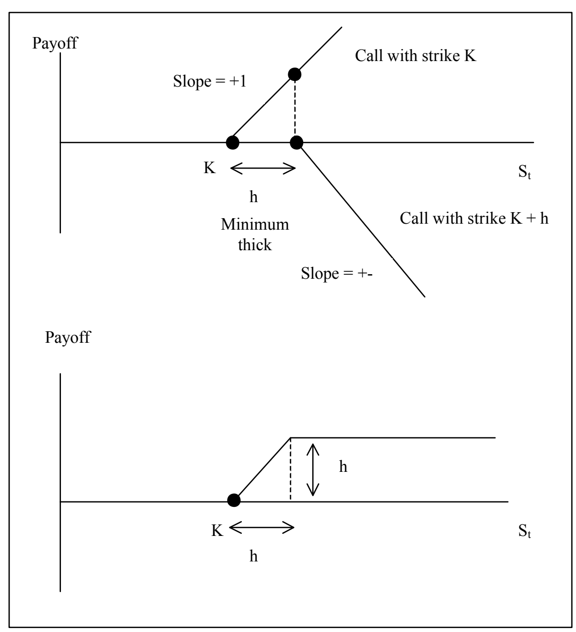


FIGURE 10-17

That is to say, the price of the binary is, in fact, the partial derivative of a vanilla call with respect to the strike price K. if all Black-Scholes assumptions hold, we can take this partial derivative analytically, and obtain

$$C^{bin}(t) = \frac{\partial C^{K}(t)}{\partial K} = e^{-r(T-t)} N(d_2) \qquad (29)$$

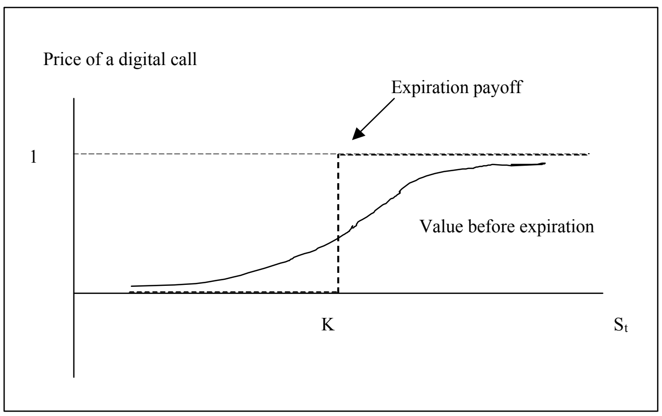
Where d_2 is, as usual,

$$d_2 = \frac{\text{Log}(S_t/K) + r(T-t) - 1/2\, \sigma^2 (T-t)}{\sigma \sqrt{(T-t)}} \qquad (30)$$

s being the constant percentage volatility of S_t, and, r being the constant risk-free spot rate.

This last result shows an interesting similarity between binary option prices and vanilla option deltas. In Chapter 9 we showed that a vanilla call's delta is given by

$$\text{Delta} = \frac{\partial C(t)}{\partial K} = N(d_1) \qquad (31)$$



Here we see that the price of the binary has a similar form. Also, it has a shape similar to that of a probability distribution:

$$C^{bin}(t) = e^{-r(T-t)} N(d_2) = e^{-r(T-t)} \int_{-\infty}^{\frac{\log(S_t/K) + r(T-t) - 1/2\, \sigma^2 (T-t)}{\sigma\sqrt{(T-t)}}} \frac{1}{\sqrt{2\pi}} e^{-1/2\, u^2} du \qquad (32)$$

This permits us to draw a graph of the binary price, $C^{bin}(t)$. Under the Black-Scholes assumptions, it is clear that this price will be as indicated in Figure 10-18.

4.1.4. Time Value of Binaries

We can use the previous result to obtain convexity characteristics of the binary option shown in Figure 10-16. The deep out-of-the-money binary will have a positive price close to zero. This price will increase and will be around ½ when the option becomes at-the-money. On the other hand, an in-the-money binary will have a price less than one, but approaching it as S_t gets larger and larger. This means that the time value of a European in-the-money binary is negative. The $C^{bin}(t)$ will never exceed 1 (or R), since a trader would never pay more than $1 in order to get a chance of earning $1 at T.

From this figure we see that a market maker who buys the binary call will be long volatility if the binary is out-of-the-money, but will be short volatility, if the binary option is in-the-money.

An ATM binary will be neutral toward volatility. This is because, in the case of an in-the-money option, the curvature of the $C^{k+h}(t)$ will dominate the curvature of the $C^{K}(t)$, and the binary will have a concave pricing function. The reverse is true if the binary is out-of-the-money.

To summarize, we see that the price of a binary is similar to the delta of a vanilla option. This implies that the delta of the binary looks like the gamma of a vanilla option. This logic tells us that the gamma of a binary looks like in Figure 10-19, and is similar to the third partial with respect to S_t of the vanilla option.

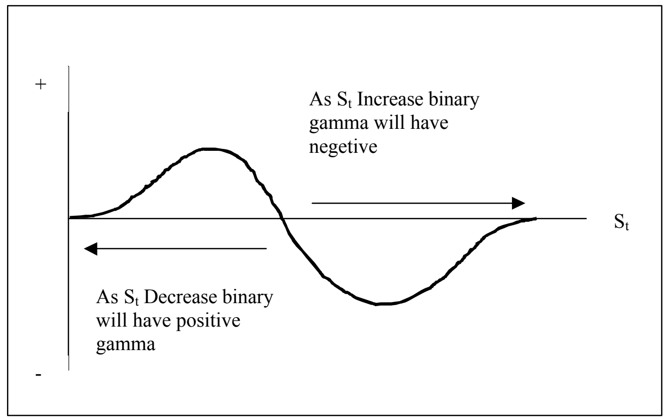


FIGURE 10-19

4.1.5. Uses of the Binary

A range option is constructed using binary puts and calls with the same payoff. This option has a payoff depending on whether the S_t remains within the range $[H^{min}, H^{max}]$ or not. Thus consider the portfolio

Range option = {Long H^{min} – Binary call, Short H^{max} – Binary call} (33)

The time-T payoff of this range option is shown in Figure 10-20. it is clear that we can use binary options to generate other, more complicated, range structures.

The expiration payoff denoted by $C^{range}(T)$ of such a structure will be given by,

$$C^{range}(T) = \begin{cases} R & \text{If } H^{min} < S_u < H^{max} \quad u \in [t, T] \\ 0 & \text{Otherwise} \end{cases} \qquad (34)$$

Thus, in this case, the option pays a constant amount R if S_u is range-bound during the whole line of the option, otherwise the option pays nothing. The following example illustrates the use of such binaries.

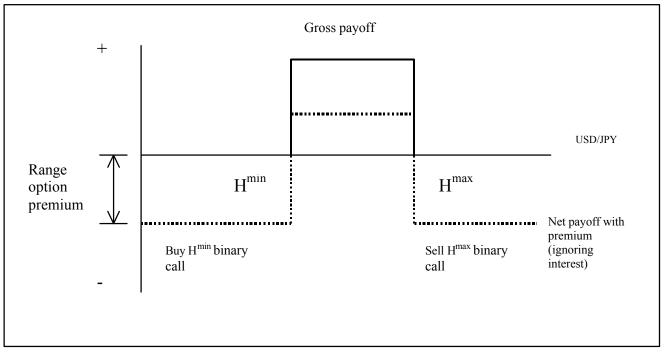


FIGURE 10-20

Example

Japanese exporters last week were snapping up one-to three-month Japanese yen/U.S. dollar binary options, struck within a JPY 114-119 range, betting that the yen will remain bound within that range. Buyers of the options get a predetermined payout if the yen trades within the range, but forfeit a principal if it touches either barrier during the life of the option. The strategy is similar to buying a yen strangle, although the down side is copped. (Based on an article in Derivatives Week)

Figure 10-20 illustrates the long binary options mentioned in the example. Looked at form the angle of yen, the binary options have similarities to selling dollar strangles. Such range structures are also called “double no touch” options.

4.2. Barrier Options

Barrier options are inexpensive instruments of applying vanilla option strategies. To create a barrier option, we basically take a vanilla counterpart, and then, add some property selected thresholds. If, during the life of the option, these thresholds are exceeded by the underlying, the option payoff will exhibit a discrete change. The option may be knocked out, or it may be knocked in, meaning that the option holder either loses the right to exercise or gains it.

Let us consider the two most common cases. We start with a European style plain vanilla option written on the underlying, S_t, with strike K, and expiration T. Next, we consider tow thresholds H^{min} and H^{max}, with $H^{min} < H^{max}$. If, during the life of the option, S_t exceeds one or both of these limits in some precise ways to be defined, then the option ceases to exist. Such instruments are called knock-out options. Two examples are shown in Figure 10-21. The lower part of the diagram is a knock-out call. If, during the life of the option, we observe the event

$S_u < H^{min}$ u € [t, T] (35)

Then the option ceases to exist. In fact, this option is down-and-out. The upper part of the diagram displays an up-and-out put, which ceases to exist if the event is observed.

$H^{max} < S_u$ u € [t, T] (36)

An option can, also, come into existence after some barrier is hit. We then call it a knock-in option. A knock-in put is shown in figure 10-22. In this section, we will discuss a H knock-out call and an H Knock-in call with the same strike K. These barrier options have the characteristic that when they knock in or out, they will be out-of-the-money. Barrier options with positive intrinsic value at knock-in and out are not dealt with. (for these, see James (2003)).

4.2.1 A Contractual Equation

We can obtain a contractual equation for barrier options and the corresponding vanilla options. Consider two European – style barrier options with the same strike K. The underlying risk is S_t, and, for simplicity, suppose all Black-Scholes assumptions are satisfied. The first option, a knock-out call, whose premium is denoted by $C^0(t)$, has the standard payoff if the S_t never touches, or falls below the barrier H. The premium of the second option, a knock-in call, is denoted by $C^I(t)$. It entitles its holder to the standard payoff of a vanilla call with strike K, only if S_t does fall below the barrier H. These payoffs are shown in Figure 10-23. In each case, h is such that, when the option knocks in or out, this occurs in a region with zero intrinsic value.

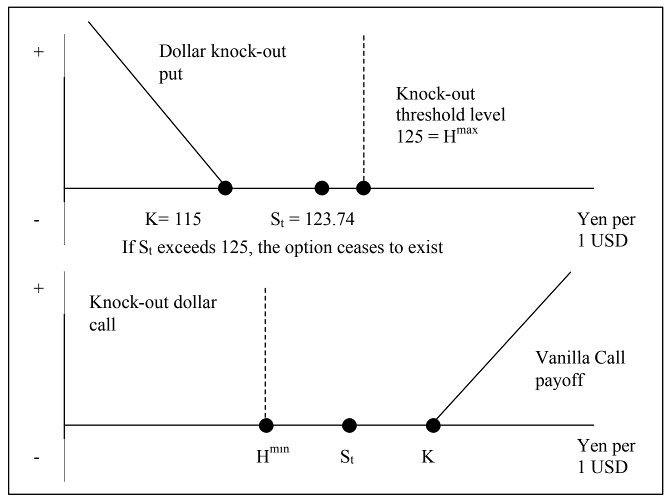


FIGURE 10-21

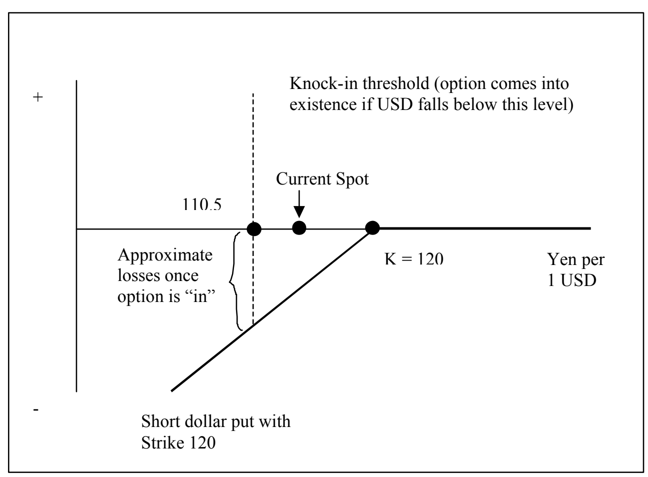


FIGURE 10-22

Now consider the following logic that will lead to a contractual equation.

1. Start with the case where S_t is below the barrier, $S_t < H$. Here, the S_t is already below the threshold H. So, the knock-out call is already worthless, while the opposite is true for the knock-in call. The knock-in is in, and the option holder has already earned the right to a standard vanilla call payoff. This means that for all $S_t < H$, the knock-in call has the same value as a vanilla call. These observations mean

For $S_t < H$ knock-in + knock-out = Vanilla call (37)

= Knock-in

The knock-out is worthless for this range.

2. Now suppose S_t is initially above the barrier, H. There are two possibilities during the life of the barrier options. S_t either stays above H, or falls below H. One and only one of these events will happen during [t, T]. this means that, if we buy the knock-in call simultaneously with a knock-out call, we guarantee access to the payoff of a vanilla call. In other words,

For $H < S_t$ Knock-in + Knock-out = Vanilla Call (39)

Putting these two payoff regions together, we obtain the contractual equation:

Vanilla Call strike K.	=	Knock-in K-Call with barrier H	+	Knock-out K-call with barrier H

From here we can obtain the pricing formulas of the knock-in and knock-out barriers. In fact, determining the pricing function of only one of these barriers is sufficient to determine the price of the other. In Chapter 8, we provided a pricing formula for the knock-out barrier where the underlying satisfied the Black-Scholes assumptions. The formula was given by

$$C^0(t) = C(t) - j(t) \quad (40)$$

Where

$$J(t) = S_t \left(H/S_t\right)^{\frac{2(r-\frac{1}{2}\sigma^2)}{\sigma^2}+2} N(c_1) - K e^{-r(T-t)} \left(H/S_t\right)^{\frac{2(r-\frac{1}{2}\sigma^2)}{\sigma^2}} N(c_2) \quad (41)$$

where

$$c_{1,2} = \frac{\ln \frac{H^2}{S_t K} + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad (42)$$

The C(t) is the value of the vanilla call given by the standard Black-Scholes formula, and the J(t) is the discount that needs to be applied because the option may disappear if S_t falls below h during [t, T].

But we now know from the contractual equation that a long knock-in and a long knock-out call with the same strike K and threshold H, is equivalent to a vanilla call:

$$C^0(t) + C^1(t) = C(t) \quad (43)$$

Using equation (40) with this gives the formula for the knock-in price as

$$C_1(t) = j(t) \quad (44)$$

Thus, the expressions in (41) – (43) provide the necessary pricing formulas for barrier options that knock-out and in, when they are out-of-the-money under the Black-Scholes assumptions.

It is interesting to note that when S_t touches the barrier,

$$S_t = H \quad (45)$$

The formula for J(t) becomes

$$J(t) = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \quad (46)$$

That is to say, the value of $C^0(t)$ will be zero. All these permit plotting the Knock-out call option price as in Figure 10-23. We see that the knock-out is higher everywhere and is discontinuous at H.

Finally, Figure 10-23 shows the pricing function of the knock-in. To get this graph, all we need to do is subtract $C_0(t)$ from C(t), in the upper part of Figure 10 -23. The reader may wonder why the knock-in call gets cheaper as S_t moves to the right of K. After all, doesn't the call become more in-the-money? The answer is no, because as long as $H < S_t$ the holder of the knock-in does not have access to the vanilla payoff yet. In other words, as S_t the holder of the knock-in does not have access to the vanilla payoff yet. In other words, as S_t the holder of the knock-in does not have access to the vanilla payoff yet. In other words, as S_t moves rightward, the chances that the knock-in call holder will end up with a vanilla option are going down.

4.2.2. Some uses of Barrier Options

Barrier options are quite liquid, especially in FX markets. The following examples discuss the payoff diagrams associated with barrier options.

The next example illustrates another way knock-ins can be used in currency markets. Figures 10-22 to 10-24 these cases.

Example

U.S. dollar puts (yen calls) were well bid last week. Demand is coming from stop-loss trading on the back of exotic knock-in structures. At the end of December some players were seen selling one-month dollar puts struck at JPY 119 which knock-in at JPY109.30. As the yen moved toward that level early last week, those players rushed to buy cover.

Hedge funds were not the only customers looking for cover. Demand for short-term dollar puts was widely seen. "People are still short yen," said a trader. "The risk reversal is four points in favor of the dollar put, which is as high as I have ever seen it" (Based on an article in Derivatives Week).

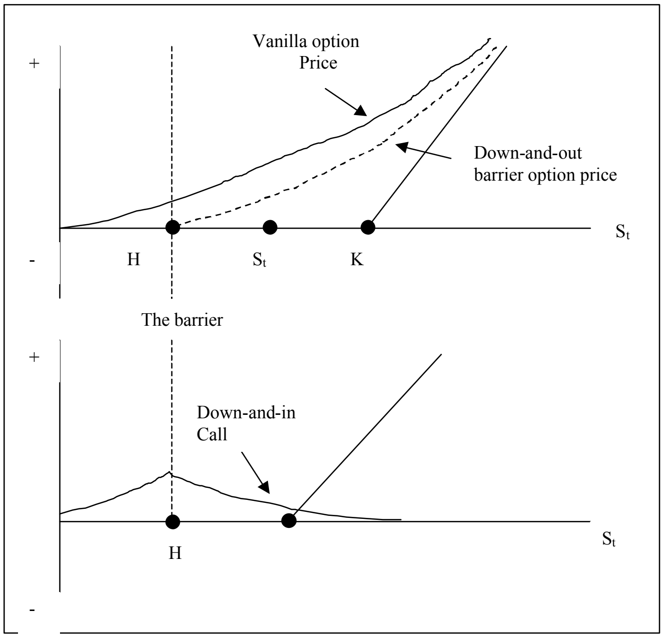


FIGURE 10-23

According to the example, a s the dollar fell toward 110.6 yen, the hedge funds who had sold knock-in options were suddenly facing the possibility that these options would come into existence, and that they would lose money. As a result, the funds started to cover their positions by buying out-of-the-money puts. This is a good illustration of new risks often associated with exotic structures. The changes during infinitesimal intervals in mark-to-market values of barrier options can be discrete instead of "gradual."

The next example concerning barrier options involves a more complex structure. The barrier may in fact relate to a different risk than the option's underlying. The example shows how barrier options can be used by the airline industry.

Airlines face three basic costs: labor, capital, and fuel. Labor costs can be "fixed" for long periods using wage contracts. However, both interest rate risk and fuel price risk are floating, and sudden spikes in these at any time can cause severe harm to an airline. The following example shows how airlines can hedge these tow risks using a single barrier option.

Example

Although these are slow days in the exotic option market, clients still want alternative ways to hedge cheaply, particularly if these hedges offer payouts linked to other exposures on their balance sheets. Barrier products are particularly popular. Corporate are trying to cheapen their projections by asking for knock-out options.

For example, an airline is typically exposed to both interest rate and fuel price risks. If interest rates rose above a specified level, a conventional cap would pay out, but under a barrier structure it may not if the airline is enjoying lower furl prices. Only if both rates and fuel prices are high is the option triggered. Consequently, the cost of this type of hedge is cheaper than separate options linked to individual exposures. (IFR, May 13, 1995)

The use of such barriers may lower hedging costs and may be quite convenient for businesses. The exercises at the end of the chapter contain further examples of exotic options. In the next section we discuss some of the new risks and difficulties associated with these.

4.3. New Risks

Exotic options are often inexpensive and convenient, but they carry their own risks. Risk management of exotic options books is nontrivial because there are (1) discontinuities in the respective Greeks due to the existence of thresholds, and, (2) smile effects in the implied volatility.

As the previous three chapters have shown, risk management of option books normally uses various Greeks or their modified counterparts. With threshold effects, some Greeks may not exist at the threshold. This introduces discontinuities and complicates risk management. We review some of these new issues next.

1. Barrier options may exhibit jumps in some Greeks. This is a new dimension in risk managing option books. When spot is near the threshold, barrier options Greeks may change discretely even for small movements in the underlying. These extreme changes in sensitively factors make the respective delta, gamma, and vega more complicated tools to use in measuring and managing underlying risks.
2. Barrier options are path dependent. For example, the threshold may be relevant at each time point until the option expires or until the barrier is hit. This makes Monte Carlo pricing and risk managing techniques more delicate and more costly. Also, near the thresholds the spot may need further simulated trajectories and this may also be costly.
3. Barrier option hedging using vanilla and digital options may be more difficult and may be strongly influenced by smile effects.

We will not discuss these risk management and hedging issues related to exotic options in this book. However, smile effects will be dealt with in Chapters 15.

4. Quoting Conventions

Quoting conventions in option markets may be very complicated. Given that market makers look at options as instruments of volatility, they often prefer quoting volatility directly, rather than a cash value for the option. These quotes can be very confusing at times. The best way to study them is to consider the case of risk reversals. Risk reversal quotes illustrate the role played by volatility, and show explicitly the existence of skewness in the volatility smile, an important empirical observation that will be dealt with separately in Chapter 15.

One of the examples concerning risk reversals presented earlier contained the following statement:

The one-month risk reversal jumped to 0.81 in favor of euro calls Wednesday from 0.2 two weeks ago.

It is not straightforward to interpret such statements. We conduct the discussion using the Euro/dollar exchange rate as the underlying risk. Consider the dollar calls represented in

Figure 10-24a, where it is assumed that the spot is trading at 0.95, and that the option is ATM. In the same figure, we also show a 25 – delta call. Similarly, figure 10-24b shows an ATM dollar put and a 25 – delta put, which will be out-of-the-money. All these options are supposed to be plain vanilla and European style.

Now consider the following quotes for two different 25-delta USD risk reversals:

Example 1: "flat/0.3 USD call bid" (47)

Example 2: "0.3/0.6 USD call bid" (48)

The interpretation of such bid-ask spreads is not straightforward. The numbers in the quotes do not relate to dollar figures, but, to volatilities. In simple terms, the number to the right of the slash is the volatility spread the market maker is willing to receive for selling the risk reversal position and the number to the left is the volatility spread he is willing to pay for the position.

The numbers to the right are related to the sale by the market maker of the 25-delta USD call and simultaneously the purchase of a 25 – delta USD put, which, from a client's point of view is the risk reversal shown in Figure 10-25a. Note that, for the client, this situation is associated with "dollar strength." If the market maker sells this risk reversal, he will be short this position.

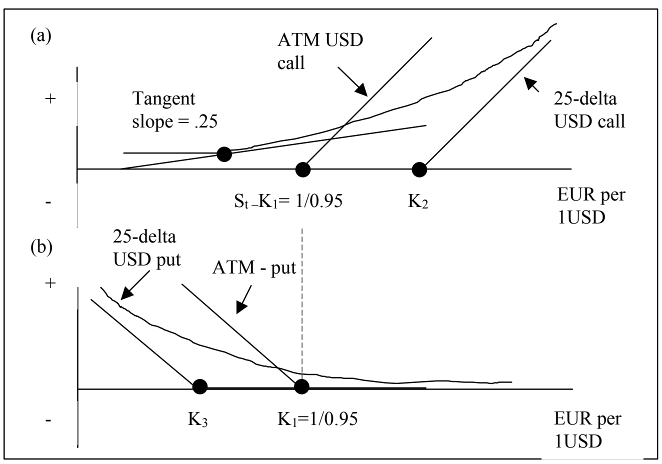


FIGURE 10-24

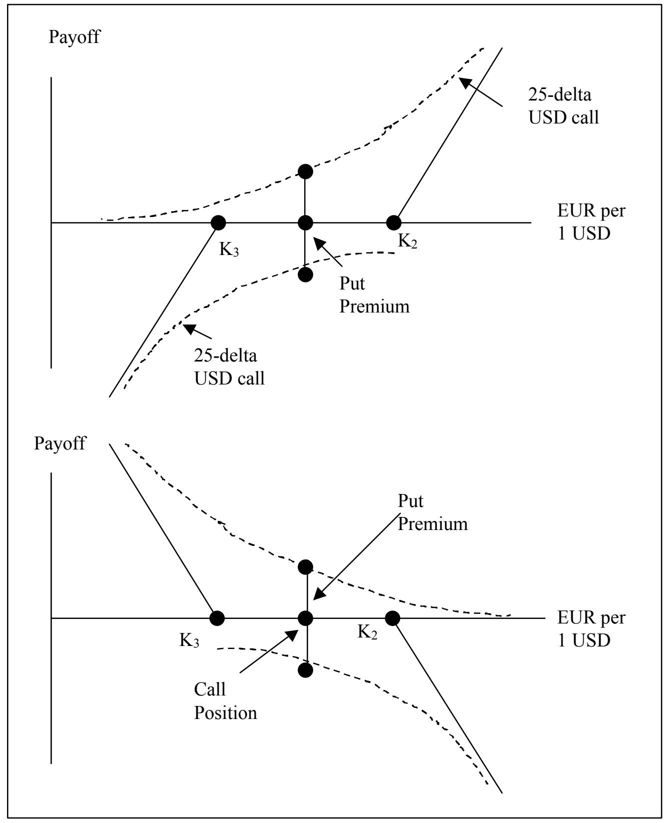


FIGURE 10-25

The numbers to the left of the slash correspond to the purchase of a 25-delta USD call and the sale of a 25-delta USD put, which is shown in figure 10-25b. This outcome, when in demand, is associated with "dollar weakness."

Example 1

Now consider the interpretation of the numerical values in the first example:

Example 1: "flat/0.3 USD call bid" (49)

The left-side in this quote is "flat." This means that the purchase of the 25-delta USD call, and a simultaneous sale of the 25-delta USD put, would be done at the same volatilities. A client who sells this to the market maker pays or receives nothing extra and the deal has "zero cost." In other words, the two sides would agree on a single volatility and then plug this same number into the Black-Scholes formula to obtain the cost of the put and the cost of the call. The right-hand number in the quote shows a bias. It means that the market maker is willing to sell the 25-delta USD call, and buy the 25-delta USD put, only if he can earn 0.3 volatility points net. This implies that the volatility number used in the sale of USD call will be 0.3 points higher than the volatility used for the 25-delta USD put. The market maker thinks that there is a "bias" in the market in favor of dollar strength; hence, the client who purchases this risk reversal will incur a net cost.

Example 2

The second quote given by

Example 2: "0.3/0.6 USD call bid" (50)

is more complicated to handle, although the interpretation of the 0.6 is similar to the first example. With this number, the market maker is announcing that he or she needs to receive 0.6 volatility points net if a client wants to bet on the dollar strength.

However, the left-hand element of the quote is not "flat" anymore but is a positive 0.3. This implies that the bias in the market, in favor of dollar strength is so large, and so many clients demand this long position that, now the market is wiling to pay net 0.3 volatility points when buying the 25-delta call and selling the 25-delta put.

Thus, in risk reversal quotes, the left-hand number is a volatility spread that the market maker is wiling to pay, and the second number is a volatility spread the market maker would like to earn. In each case, to see how much the underlying option would cost, market participants have to agree on some base volatility and then, using it as a benchmark, bring in the volatility spreads.

6. Real-World Complications

Actual implementation of the synthetic payoff structures discussed in this chapter requires dealing with several real-world imperfections. First of all, it must be remembered that these positions are shown at expiration, and that they are piecewise linear. In real life, payoff diagrams may contain several convexities, which is an equivalent term for nonlinear payoffs. We will review these briefly.

6.1 The Role of the Volatility smile

The existence of volatility smile has especially strong effects on pricing and hedging of exotic options. If a volatility smile exists, the implied volatility becomes a function of the strike price K. Then, the expression that gave the binary option price in (29) – (30) has to be modified to

$$C^{bin}(t) = \lim_{h \to 0} \frac{C^k(t) - C^{k+h}(t)}{h} \qquad (51)$$

$$= \frac{\partial C^k(t)}{\partial K} + \frac{\partial C^k(t)\ \partial \sigma(K)}{\partial \sigma(K)\ \partial K} \qquad (52)$$

The resulting formulas and the analogy to plain vanilla deltas will change. These types of modifications have to be applied to hedging and synthetically creating barrier options as well.

6.2. Existence of Position Limits

At time t before expiration, an option's value depends on many variables other than the underlying x_t. The volatility of x_t and the risk-free interest rate r_t are two random variables that affect all the positions discussed for t < T. This is expressed in the Black-Scholes formula for the call premium of t < T:

$$C_t = C(x_t, t \mid s, r) \qquad (53)$$

Which is a function of the "parameters" r, s . At t=T this formula reduces to

$$C_T = \max[x_t - K, 0] \qquad (54)$$

Now, if he r and s are stochastic, then during the t € [0, T], the positions considered here will be subject to vega and rho risks as well. A player who is subject to limits on how much of these risks he or she can take, may have to unwind the position before T. This is especially true for positions that have vega risk. The existence of limits will change the setup of the problem since, until now, sensitivities with respect to the r and ó parameters did not enter the decision to take and maintain the positions discussed.

BINOMIAL OPTION PRICING MODELS

Objective

- After completion of this lesson you will be able to of how binomial option is important in stock options.

A useful and very popular technique for pricing a stock option involves constructing a binomial tree. This is a diagram that represents different paths that might be followed by the stock price over the life of the option. In this lesson, we will take a first look at binomial trees and their relationship to an important principle known as risk-neutral valuation. The approach we adopt here is similar to that in an important paper published by Cox, Ross, and Rubinstein in 1979.

The material in this chapter is intended to be introductory. More details on the use of numerical procedures involving binomial trees are given in Chapter 18.

10.1 A One-step Binomial Model

We start by supporting that we are interested in valuing a European call option to buy a stock for \$21 in three months. A stock price is currently \$20. We make a simplifying assumption that at the end of three months the stock price will be either \$22 or \$18. This means that the option will have one of two values at the end of the three months. If the stock price turns out to be \$22, the value of the option will be \$1; if the stock price turns out to be \$18, the value of the option will be zero. The situated is illustrated in figure 10.1.

It turns out that an elegant argument can be used to price the option in this situation. The only assumption needed is that no arbitrage opportunities exist. We set a portfolio of the stock and the option in such a way that there is no uncertainty about the value of the portfolio at the end of the three months. We then argue that, because the portfolio has no

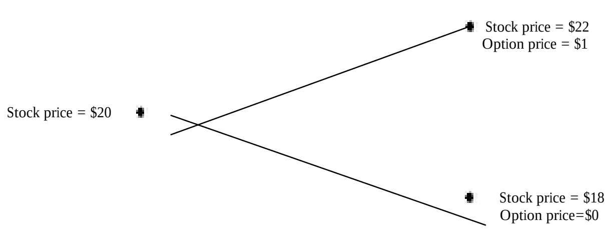


Figure 10.1 Stock Price Movements In Numeriacl Example

Risk, the return it earns must equal the risk-free interest rate. This enables us to work out the cost of setting up the portfolio and therefore the option's price. Because there are two securities (the stock and the stock option) and only two possible outcomes, it is always possible to set up the risk less portfolio.

Consider a portfolio consisting of a long position in Δ shares of the stock and a short position in one call option. We calculate the value of that makes the portfolio riskless. If the stock price moves up from \$20 to \$22, the value of the share is 22 and the value of the option is 1, so that the total value of the portfolio is 22 - 1. if the stock price moves down from \$20 to \$18, the value of the shares is 18 and the value of the option is zero, so that the total value of the portfolio is 18. The portfolio is risk less if the value of is chosen so that the final value of the portfolio is the same for both alternatives. This means

22 - 1 = 18

or

= 0.25

A riskless portfolio is therefore

Long: 0.25 shares

Short: 1 option

If the stock price moves up to \$22, the value of the portfolio is

$22 \times 0.25 - 1 = 4.5$

If the sock price moves down to \$18, the value of the portfolio is

$18 \times 0.25 = 4.5$

Regardless of whether the sock price moves up or down, the value of the portfolio is always 4.5 at the end of the life of the option.

Riskless portfolios must, in the absence of arbitrage opportunities, earn the risk-free rate of interest. Suppose that in this case the risk-free rate is 12% per annum. It follows that the value of the portfolio today must be the present value of 4.5, or

$4.5e^{-0.12 \times 3/12} = 4.367$

The value of the stock price today is known to be \$20. Suppose the option price is denoted by f. The value of the portfolio today is

$20 \times 0.25 - f = 5 - f$

It follows that

$5 - f = 4.367$

Or

$f = 0.6333$

This shows that, in the absence of arbitrage opportunities, the current value of the option must be 0.6333. If the value of option were more than 0.6333, the portfolio would cost less than 4.367 to set up and would earn more than the risk-free rate. If the value of the option were less than 0.6333, shorting the portfolio would provide a way of borrowing money at less than the risk-free rate.

A Generalization

We can generalize the argument just presented by considering a stock whose price is S_0 and an option on the stock whose current price is f. We suppose that the option lasts for time T and that during the life of the option the stock either move up from S0 to a new level S_0u or down from S_0 to a new level S_0 d, where u >1 and d<1. The proportional increase in the stock price when there is an up movement is u - 1; the proportional decrease when there is a down movement is 1- d. if the stock price moves to S_0 d, we suppose that the payoff from the option is f_d. The situation is illustrated in Figure 10.2.

As before, we imagine a portfolio consisting of a long position in Δ shares and a short position in one option. We calculate the value of that makes the portfolio risk less. If there is an up movement in the stock price, the value of the portfolio at the end of the option is

$S_0u - f_u$

If there is a down movement in the stock price, the values becomes

$S_0d - f_d$

The two are equal when

$S_0u - f_u = S_0d - f_d$

Or

$$\Delta = \frac{f_u - f_d}{S_0u - S_0d} \qquad (10.1)$$

In this case, the portfolio is risk less and must earn the risk-free interest rate. Equation (10.1) shows that Δ is the ratio of the change in the option price to the change in the stock price as we move between the nodes.

If we denote the risk-free interest rate by r, the present value of the portfolio is

$(S_0u - f_u)\, e^{-rT}$

The cost of setting up the portfolio is

$S_0u - f$

It follows that

$S_0u - f = (S_0u - f_u)\, e^{-rT}$

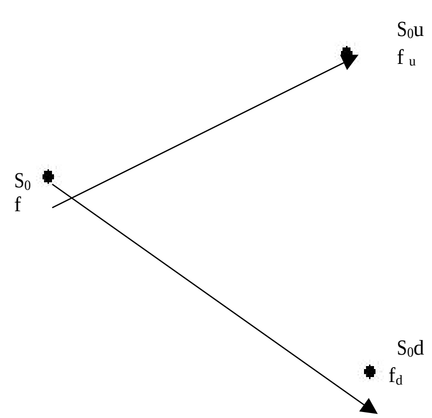


Figure 10.2 : Stock and Option Prices In a General One-step Tree

Or

$$F = S_0 \Delta - (S_0u - f_u)\, e^{-rT}$$

Substituting for from equation (10.1) and simplifying reduces this equation to

$$F = e^{-rT}\,[p f_u + (1- p) f_d] \qquad (10.2)$$

Where

$$P = \frac{e^{-rT} - d}{u - d} \qquad (10.3)$$

Equations (10.2) and (10.3) enable an option to be priced using a one-step binomial model.

In this numerical example considered previously (see Figure 10.1), u =1.1, d=0.9, r = 0.12, T= 0.25, f_u =1 and f_d =0. From equation (10.3),

$$P = \frac{e^{-0.12 \times 3/12} - 0.9}{1.1 - 0.9} = 0.6523$$

And, from equation (10.2),

$F = e^{-0.12\times 3/12}\,(0.6523 \times 1 + 0.3477 \times 0) = 0.633$

The result agrees with the answer obtained earlier in this section.

Irrelevance of the Stock's Expected Return

The option-pricing formula in equation (10.2) does not involve the probabilities of the stock price moving up or down. For example, we get the same option price when the probability of an upward movement is 0.5 as we do when it is 0.9. This is surprising and seems counterintuitive. It is natural to assume that, as that probability of an upward movement in the stock price increases, the value of a call option on the stock increases and the value of a put option on the stock decreases. This is not the case.

The key reason is that we are not valuing the option in absolute terms. We are calculating its value in terms of the price of the underlying stock. The probabilities of future up or down movements are already incorporated into the price of the stock. It turns out that we do not need to take them into account again when valuing the option in terms of the stock price.

10.2 Risk- Neutral Valuation

Although we do not need to make any assumptions about the probabilities of up and down movements in order to derive equation (10.2), it is natural to interpret the variable p in equation (10.2) as the probability of an up movement in the stock price. The variable 1— p is then the probability of a down movement, and the expression

$P\, f_u + (1— p)\, f_d$

Is the expected payoff from the option? With this interpretation of p, equation (10.2) then states that the value of the option today is tis expected future value discounted at the risk-free rate.

We now investigate the expected return the stock when the probability of an up movement is assumed to be p. the expected stock price, $E(S_T)$, at time T is given by

$$E(S_T) = pS_o u + (1—p)\, S_o d$$

Or $$E(S_T) = pS_o\,(u —d) + S_o d$$

Substituting from equation (10.3) for p, we obtain

$$E(S_T) = S_o e^{rT} \qquad (10.4)$$

Showing that the stock price grows on average at the risk-free rate. Setting the probability of the up movement equal to p is therefore equivalent that the return on the stock equals the risk-free rate.

In a risk-neutral world all individuals are different to risk. In such world investors require no compensation for risk, and the expected return on all securities is the risk-free interest rate. Equation (10.4) shows that we are assuming a risk-neutral world when we set the probability of an up movement to p. equation (10.2) shows that the value of the option is its expected payoff in a risk-neutral world discounted at the risk-free rate.

This result is an example of an important general principle in option pricing known as risk-neutral valuation. The principle states that we can assume the world is risk neutral when pricing an option. The price we obtain is correct not just in a risk-neutral world but in the real world as well.

The One-step Binomial Example Revisited

We now return to the example in Figure 10.1 and illustrate that risk-neutral valuation gives the same answer as no-arbitrage arguments. In figure 10.1, the stock price is currently $20 and will move either up to $22 or down to $18 at the end of three months. The option considered is a European call option with a strike price of $21 and an expiration date in three months. The risk-free interest rate is 12% per annum.

We define p as the probability of an upward movement in the stock in a risk-neutral world. We calculate p from equation (10.3). Alternatively, we can argue that the expected return on the stock in a risk-neutral world must be the risk-free rate of 12%. This means that p must satisfy

$$22p + 18(1—p) = 20e^{0.12\times 3/12}$$

Or

$$4p = 20e^{0.12\times 3/12} — 18$$

That is p must be 0.6523

At the end of the three months, the call option has a 0.6523 probability of being worth 1 and a 0.3477 probability of being worth zero. Its expected value is therefore

$$0.6523\times 1 + 0.3477\times 0 = 0.6523$$

In a risk-neutral world this should be discounted at the risk-free rate. The value of the option today is therefore

$$0.5623\, e^{0.12\times 3/12}$$

Or $0.633. This is the same as the value obtained earlier, demonstrating that no-arbitrage arguments and risk-neutral valuation give the same answer.

Real World vs. Risk-neutral World

It should be emphasized that p is the probability of a up movement in a risk-neutral world. In general that is not same as the probability of an up movement in the real world. In our example, p=0.6523. When the probability of an up movement is 0.6523, the expected return on the stock is the risk-free rate of 12%. Suppose that in the real world the expected return on the stock is 16%. And q is the probability of an up movement in the real world. It follows that

$$22q + 18(1-q) = 20e^{0.16\times 3/12}$$

So that q= 0.7041.

The expected payoff from the option in the real world is then

$$Q\times 1 + (1-q)\times 0$$

This is 0.7041. Unfortunately it is not easy to know the correct discount rate to apply th the expected payoff in the real world. A position in a call option is risker than a position in the stock. As a result the discount rate to the payoff from a call option is greater than 16%. Without knowing the option's value, we do not know how much greater than 16% it should be. The risk-neutral valuation solves this problem. We know that in a risk-neutral world the expected return on all assets (and therefore the discount rate is to use for all expected payoffs) is the risk-free rate.

10.3 Two-step Binomial Trees

We can extend the analysis to a two-step binomial tree such as that shown in figure 10.3. Here the stock price starts at $20 and in each of two time steps may go up 10% or down by 10%. We suppose that each time step is three months long and the risk-free interest rate is 12% per annum. As before, we consider an option with a strike price of $21.

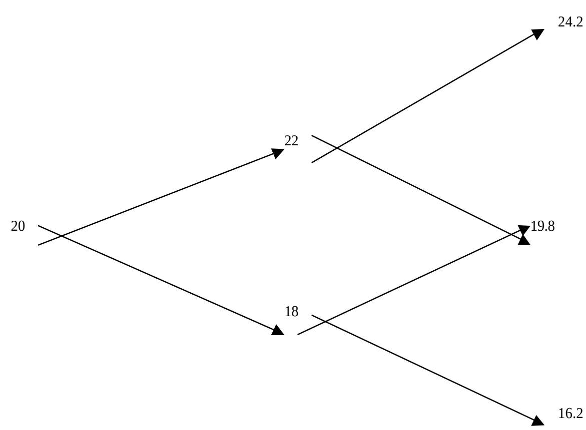
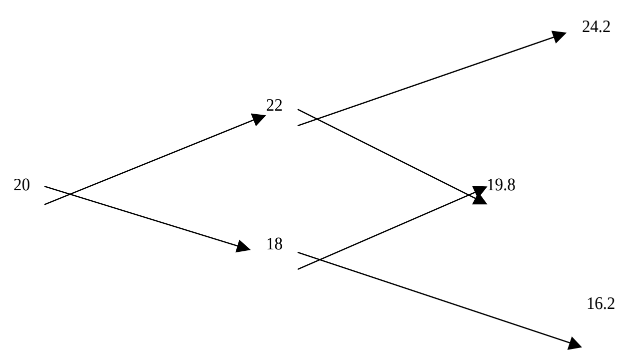


Figure 10.3 Stock Prices In A Two-step Tree



10.4 Stock and Option Prices in a Two - Step Tree. The Upper Number at Each Node is Stock Price, The Lower Number is the Option Price

The objective of the analysis is to calculate the option price at the initial node of the tree. This can be done by repeatedly applying the principles established earlier in the chapter. Figure 10.4 shows the same tree as figure 10.3 but with both the stock price and the opinion price at each node. (The stock price is the upper number and the option price is the lower number.) The opinion prices at the final nodes of the tree are easily calculated. They are the payoffs from the option. At node D the stock price is 24.2 and the option price is 24.2—21 = 3.2; at nodes E & F the option is out of the money and its value is zero.

At node C, the option price is zero, because node C leads to either node E or node F and at both nodes the option price is zero. We calculate the option price at node B by focusing our attention on the part o the tree shown in Figure 10.5. Using the notation introduced earlier in the chapter, u=1.1, d=0.9, r=0.12, and T=0.25, so that p=0.5623, and equation (10.2) gives the value of the option at node B as

$e^{0.12 \times 3/12}$ $(0.6523 \times 3.2 + 0.3477 \times 0) = 2.0257$

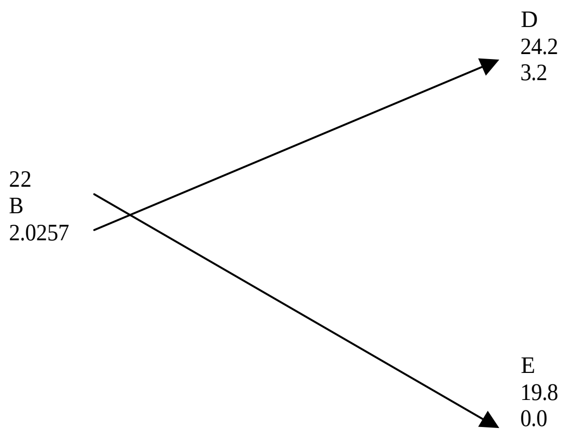


Figure 10.5 : Evaluation of Option Price at Node B

It remains for us to calculate to option price at the initial node A. We do so by focusing on the first step of the tree. We know that the value of the option at node B is 2.0257 and that at node C it is zero. Equation (10.2) therefore gices the value at node A as

$e^{0.12 \times 3/12}$ $(0.6523 \times 2.0257 + 0.3477 \times 0) = 1.2823$

The value of the option is \$1.2823

Note that this example was constructed so that u and d (the proportional up and down movements) were the same at each node of the tree and so that the time steps were of the same length. As a result, the risk-neutral probability, p, as calculated by equation (10.3) is the same at each node.

A Generalization

We can generalize the case of two time steps by considering the situation in Figure 10.6.

The stock price is initially S_0. During each time step, it either moves up to u times its initial value or moves down to d times its initial value. The notation for the value of the option is shown in the tree. (For example, after two up movements the value of the option is f_{uu}). We suppose that the risk free interest rate is r and the length of the time step is dt years.

Repeated application of equation (10.2) gives

$$f_u = e^{rdt} [p f_{uu} + (1 - p) f_{ud}] \quad (10.5)$$

$$f_d = e^{-rdt} [p f_{ud} + (1 - p) f_{dd}] \quad (10.6)$$

$$f_u = e^{-rdt} [p f_u + (1 - p) f_d] \quad (10.7)$$

Substituting from equations (10.5) and (10.6) into (10.7), we get

$$f = e^{-2rdt} [p^2 f_{uu} + 2p(1 - p) f_{ud} + (1-p)^2 f^{dd}] \quad (10.8)$$

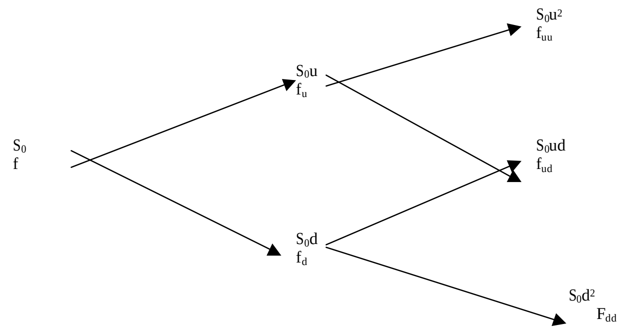


Figure 10.6 Stock and Option Prices In a General Two Step Tree

This is consistent with the principle of risk-neutral valuation mentioned earlier. The variables p^2, $2p(1-p)$ and $(1-p)^2$ are the probabilities that the upper, middle, and lower final nodes will be reached. The option price is equal to its expected payoff in a risk-neutral world discounted at the risk-free interest rate.

As we add more steps to the binomial tree, the risk-neutral valuation principle continues to hold. The option price is always equal to its expected payoff in a risk-neutral world, discounted at the risk-free interest rate.

10.4 A Put Example

The procedure described in this chapter can be used to price any derivative dependent on a stock whose price changes are binomial. Consider a two-yea European put with a strike price of \$52 in a stock whose current price is \$50. We suppose that there are two time steps of one year, and in each time step the stock price either moves up by a proportional amount of 20% or moves down by a proportional amount of 20%. We also suppose that the risk-free interest rate is 5%.

The tree is shown in figure 10.7. The value of the risk-neutral probability, p, is given by

$$P = \frac{E^{0.05 \times 1} - 0.8}{1.2 - 0.8} = 0.6282$$

The possible final stock prices are:

\$72, \$48, and \$32. In this case $f_{uu} = 0$, $f_{ud} = 4$ and $f_{dd} = 20$. From equation (10.8)

$F = E^{-2*0.05 \times 1} (0.6282^2 \times 0 + 2 \times 0.6282 \times 0.3718 \times 4 + 0.3718^2 \times 20) = 4.1923$

The value of the put is \$4.1923. This result can also be obtained using equation (10.2) and working back through the tree one step at a time. Figure 10.7 shows the intermediate option prices that are calculated.

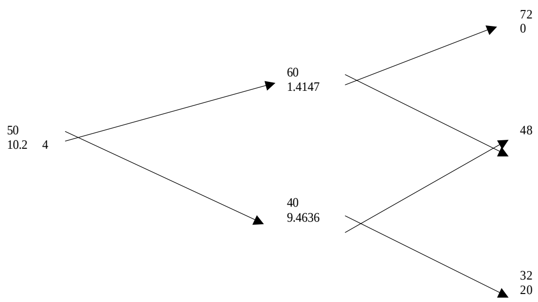


FIGURE 10.7 Use of two-step tree to value European put option. At each node, the upper number is the stock is the price and the lower number is the option price

10.4 American Options

Up to now all the options we have considered have been European. We now move on to consider how American options can be valued using a binomial tree such as that in figure 10.4 or 10.7. The procedure is to work back through the tree from the end to the beginning, testing at each node to see whether early exercise is optimal. The value of the option at the final nodes is the same as for the European options. At earlier nodes the value of the option is the greater of

1. The value given be equation (10.2)
2. The payoff from early exercise

Figure 10.8 shows how figure 10.7 is affected if the option under consideration is American rather than European. The stock prices and their probabilities are unchanged. The values for the option at the final nodes are also unchanged. At node B, equation (10.2) gives the value of the option as 1.4147, whereas the payoff from early exercise is negative (= -8). Clearly early exercise is not optimal at node B, and the value of the option at this node is 1.4147. At node C, equation (10.2) gives the value of the option as 9.4636, whereas the payoff from early exercise is 12. In this case, early exercise is optimal and the value of the option at the node is 12. at the initial node A, the value given by equation (10.2) is

$$E^{-0.05\times 1}(0.6282\times 1.4147 + 0.3718\times 12) = 5.0894$$

And the payoff from early exercise is 2. In this case early exercise is not optimal, and the value of the option is therefore $5.0894. More details on the use of binomial trees to value American options are given in chapter 18.

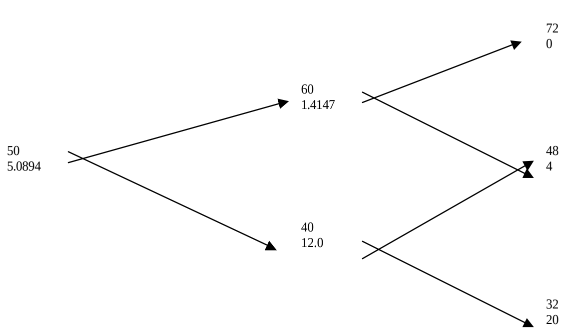


Figure 10.7 Use of two-step tree to value American put option. At each node, the upper number is the stock is the price and the lower number is the option price

10.5 Delta

At this stage it is appropriate to discuss delta, an important parameter in the pricing and hedging of options.

The delta of a stock option ois the ratio of the change in the price of the stock option to the change in the price of the underlying stock. It is the number of units of the stock we should hold for each option shorted in order to create a risk less hedge. It is same as the Δ introduced earlier in this chapter. The construction of a risk less hedge is sometime referred to as delta hedging. The delta of a call option is positive, whereas the delta of a put option is negative.

From figure 10.1, the delta we can calculate the value of the delta of the call being considered as

$$\frac{1-0}{22-18} = 0.25$$

This is because when the stock price changes from $18 to $22, the option price changes from $0 to $1. In Figure 10.4, the delta corresponding to stock price movements over the first time step is

$$\frac{1.257 - 0}{22 - 18} = 0.5064$$

The delta for stock price movements over the second time step is

$$\frac{3.2 - 0}{24.2 - 19.8} = 0.7273$$

If there is an upward movement over the first time step and

$$\frac{0 - 0}{19.8 - 16.2} = 0$$

If there is a downward movement over the first time step.

From Figure 10.7, delta is

$$\frac{1.4147 - 9.4636}{60 - 40} == -0.4024$$

at the end of the first time step and either

$$\frac{0 - 4}{72 - 48} = -0.1667$$

or

$$\frac{4 - 20}{48 - 32} = -1.0000$$

at the end of the second time step.

10.7 Matching Volatility with u and d

in practice, when constructing a binomial tree to represent the movements in a stock price, we choose the parameters u and d to match the volatility of stock price. To see how this is done, we suppose that the expected return ion a stock (in the real world) is m and its volatility is s. Figure 10.9a shows stock price movements over the first step of a binomial tree. The step is of length dt. the stock price either moves up by a proportional amount u or moves down by a proportional amount d. the probability of an up movement (in the real world) is assumed to be q.

the expected stock price at the end of the first time step is $S_o e^{u d t}$. on the tree the expected stcok price at this time is

$qS_o u + (1—q) S_o d$

in order tomatch the expected return on the stock with the tree's parameters, we must therefore have

$qS_o u + (1—q) S_o d = S_o e^{u d t}$

as we will explain in chapter 11, the volatility s of a stock price is defined so that $s\sqrt{d}$ t is the standard deviation of the return on the stock price in a short period of time of length d t. equivalently, the variance of the return is $s^2 d$ t. on the tree in figure 10.9a, the variance of the stock price return is

Qu2 + (1— q) d2 – [qu + (1— q) d]^2

In order to match the stock price volatility with the tree's parameters, we must therefore have

$$Qu2 + (1— q) d2 – [qu + (1— q) d]^2 == \sigma^2\delta t \qquad (10.10)$$

Substituting from equation 10.9 into equation (10.10), we get

$$e^{\mu\delta_t} (u + d) – ud – e^{2\mu\delta_t} == \sigma^2\delta t$$

When terms in d t^2 and higher powers of dt are ignored, one solution to this equation is

$$U = e^{\sigma\sqrt{\delta_t}} \qquad (10.11)$$

$$D = e^{\sigma\sqrt{\delta_t}} \qquad (10.12)$$

These are the values of u and d proposed by Cox, Ross, and Rubin stein (1979) for matching u and d.

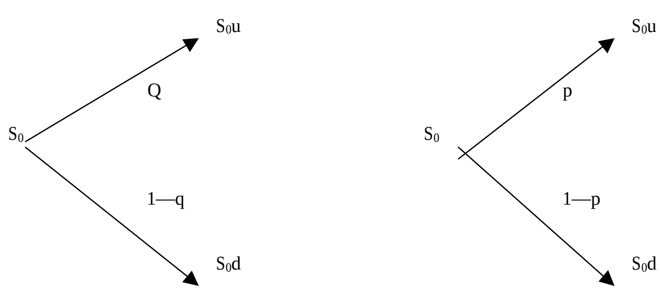


Figure 10.9 change in stock in time d t in (a) the real world and (b) the risk-neutral world

The analysis in this chapter shows that we replace the tree in figure 10.9a by the tree in Figure 10.9b where the probability of a up movement is p, and then behave as though the world is risk neutral. The variable p is given by equation (10.3) as

$$p = \frac{e^{r\delta_t} – d}{u – d}$$

It is the risk-neutral probability of an up movement. In figure 10.9b the expected stock price at the end of the time step is $S_0 e^{r d t}$ as shown in equation (10.4) the variance of the stock price return is

$$pu^2 + (1—p)d^2 – [pu + (1—p)d]^2 = [e^{r\delta_t} (u+d) – ud – e^{r\delta_t}]$$

Substituting for u and d from equation (10.11) and 10.12 we, find this equals $s^2 dt$ when terms in dt^2 and higher powers of d t are ignored.

This analysis shows that when we move from the real world to the risk-neutral world the expected return on the stock changes but its volatility remains the same (at least in the limit as d t tends to zero). This is an illustration of an important general known as GIRSANOV'S THEOREM. When we move from a world with one set of risk preferences to a world with another set of risk preferences, the expected growth rates in variables change, but their volatility remains the same. We will examine the impact of risk preferences on the behaviour of market variables in more detail . Moving from one set of risk preferences to another is something referred to as CHANGING THE MEASURE?

10.8 BINOMIAL TREES IN PRACTICE

The binomial models presented so far have been unrealistically simple. Clearly an analyst can expect to obtain only a very rough approximation to an option price by assuming that stock price movements during the life of the option consists of one or two binomial steps.

When binomial tress are used in practice, the life of the option is typically divided into 30 or more times steps of length d t. in each time step there is a binomial stock price movement. With 30 times steps, this means that 31 terminal stock prices and 2^{30}, or about 1 billion, possible stock price paths are considered.

The parameters u and d are chosen to match the clock price volatility. A popular way of doing this is by setting

$$U = e^{\sigma\sqrt{\delta_t}} \quad \text{and} \quad d = e^{-\sigma\sqrt{\delta_t}}$$

As indicated in the previous section. The complete set of equations defining the tree is then

$$U = e^{\sigma\sqrt{\delta_t}} \quad \text{and} \quad d = e^{-\sigma\sqrt{\delta_t}}$$

A further discussion of these formulas and the practical issues involved in the construction and use of binomial trees can be discussed later. Deriver Gem provides a way of valuing options with between 2& 500 times steps.

Option Pricing A Simplified Approach

John C. Cox, Stephen A. Ross and Mark Rubinstein

Introduction

An option is a security which gives its owner the right to trade in a fixed number of shares of a specified common stock at a fixed price at any time on or before a given date. The act of making this transaction is referred to as exercising the option. The fixed price is termed the striking price, and the given date, the expiration date. A call option gives the right to buy the shares; a put option gives the right to sell the shares.

Options have been traded for centuries, but they remained relatively obscure financial instruments until the introduction of a listed options exchange in 1973. Since then, options trading has enjoyed an expansion unprecedented in American securities markets.

Option pricing theory has a long and illustrious history, but it also underwent a revolutionary change in 1973. At that time, Fischer Black and

* Our best thanks go to William Sharpe, who first suggested to us the advantages of the discrete-time approach to option pricing developed here. We are also grateful to our students over the past several years. Their favorable reactions to this way of presenting things encouraged us to write this article. We have received support from the National Science Foundation under Grants Nos. SOC-77-18087 and SOC-77-22301.

Myron Scholes presented the first completely satisfactory equilibrium option pricing model. In the same year, Robert Merton extended their model in several important ways. These path-.breaking articles have formed the basis for many subsequent academic studies.

As these studies have shown, option pricing theory is relevant to almost every area of finance. For example, virtually all corporate securities can be interpreted as portfolios of puts and calls on the assets of the firm. [1] Indeed, the theory applies to a very general class of economic problems - the valuation of contracts where the outcome to each party depends on a quantifiable uncertain future event.

Unfortunately, the mathematical tools employed in the Black-Scholes and Merton articles are quite advanced and have tended to obscure the underly-ing economics. However, thanks to a suggestion by William Sharpe, it is possible to derive the same results using only elementary mathematics.[2]

In this article we will present a simple discrete-time option pricing formula. The fundamental economic principles of option valuation by arbitrage methods are particularly clear in this setting. Sections 2 and 3 illustrate and develop this model for a call option on a stock which pays no dividends. Section 4 shows exactly how the model can be used to lock in pure arbitrage profits if the market price of an option differs from the value given by the model, In section 5, we will show that our approach includes the Black-s school model as a special limiting case. By taking the limits in a different way, we will also obtain the Cox-Ross (1975) jump process model as another special case,

Other more general option pricing problems often seem immune to reduction to a simple formula. Instead, numerical procedures must be employed to value these more complex options. Michael Brennan and Eduardo Schwartz (1977) have provided many interesting results along these fines. However, their techniques are rather complicated and are not directly related to the economic structure *of* the problem. Our formulation, by its very construction, leads to art alternative numerical procedure which is both simpler, and for many purposes, computationally more efficient.

Section 6 introduces these numerical procedures and extends the model to include puts and calls on stocks which pay dividends. Section 7 concludes the paper by showing how the model can be generalized in other important ways and discussing its essential role in valuation by arbitrage methods.

[1] To take an elementary case, consider a firm with a single liability of a homogeneous class of pure discount bonds. The stockholders then have a 'call' on the assets of the firm which they can choose to exercise at the maturity date of the debt by paying its principal to the bondholders. In turn, the bonds can be interpreted as a portfolio containing a default-free loan with the same face value u the bonds and a short position in a put on the assets of the firm. .

[2] Sharpe (1918) has partially developed this approach to option pricing in his excellent new book, Investments. Rendleman and Bartter (1978) have recently independently discovered a similar formulation of the option pricing problem.

2. The Basic Idea

Suppose the current price of a stock is S = $50, and at the end of a period of time. its price must be either S* = $25 or S* = $100. A call on the stock is available with a striking price of *K* =$50, expiring at the end of the period.[3] It is also possible to borrow and lend at a 25 % rate of interest. The one piece of information left unfurnished is the current value of the call, C. However. if risk less profitable arbitrage is not possible, we can deduce from the given information *alone* what the value of the call *must* be!

Consider forming the following levered hedge:

a. Write 3 calls *C* each,
b. Buy 2 shares at $50 each, and
c. Borrow $40 at 25%, to be paid back at the end of the period.

Table 1 gives the return from this hedge for each possible level of the stock price at expiration. Regardless of the outcome, the hedge exactly breaks even on the expiration date. Therefore, to prevent profitable risk less arbitrage, its current cost must be zero; that is,

3C -100+40=0.

The current value of the call must then be C = $20.

Table 1 : Arbitrage table illustrating the formation of a riskless hedge.

	Present Date	Expiration Date	
		$S^* = \$25$	$S^* = 100$
Write 3 calls	3C	-	-150
Buy 2 shares	-100	50	200
Borrow	40	-50	-50
Total		-	-

If the call were not priced at $20, a sure profit would be possible. In particular, if C = $25, the above hedge would yield a current cash inflow of $15 and would experience no further gain or loss in the future. On the other hand, if C = $15, then the same thing could be accomplished by buying 3 calls selling short 2 shares, and lending $40.

[3]To keep matters simple, assume for now that the stock will pay no cash dividends during the life of the call. We also ignore transaction costs margin requirements and taxes.

Table 1 can be interpreted as demonstrating that *an appropriately levered position in stock will replicate the future returns of a call.* That is, if we buy shares and borrow against them in the right proportion, we can, in effect, duplicate a pure position in calls. In view of this, it should seem less surprising that all we needed to determine the *exact* value of the call was its *striking price, underlying stock price, range of movement in the underlying stock price, and the rate of interest.* What may seem more incredible is what we do not need to know: among other things, *we do not need to know the probability that the stock price will rise or fall.* Bulls and bears must agree on the value of the call, relative to its underlying stock price!

This example is very simple, but it shows several essential features of option pricing. And we will soon see that it is not as unrealistic as its seems.

3. The Binomial Option Pricing Formula

In this section, we will develop the framework illustrated in the example into a complete valuation method. We begin by assuming that the stock price follows a multiplicative binomial process over discrete periods. The rate of return on the stock over each period can have two possible values: $u - 1$ with probability q, or d - 1 with probability 1 - q. Thus, if the current stock price is S, the stock price at the end of the period will be either uS or dS. We can represent this movement with the following diagram:

uS with probability q,

S

dS with probability 1 - q.

We also assume that the interest rate is constant. Individuals may borrow or lend as much as they wish at this rate. To focus on the basic issues, we will continue to assume that there are no taxes, transaction costs, or margin requirements. Hence, individuals are allowed to sell short any security and receive full use of the proceeds.[4]

Letting r denote one plus the risk less interest rate over one period, we require $u > r > d$. If these inequalities did not hold, there would be profitable risk less arbitrage opportunities involving only the stock and riskless borrow-ing and lending.[5]

To see how to value a call on this stock, we start with the simplest situation: the expiration date is just one period away. Let C be the current value of the call, C_u be its value at the end of the period if the stock price

[4]Of course, restitution is required for payouts made to securities held short.

[5]We will ignore the uninteresting special case where q is zero or one and $u=d=r$.

goes to uS. and C_d be its value at the end of the period if the stock price goes to dS. Since there is now only one period remaining in the life of the call. we know that the terms of its contract and a rational exercise policy imply that Cu=max[0,dS-K] and Cd=max[0,dS-K]. Therefore.

C_u = max[O, uS -K] with probability q,

C

C_d=max[O.dS-K] with probability l-q.

Suppose we form a portfolio containing • shares of stock and the dollar amount B in risk less bonds.6 This will cost • S + B. At the end of the period, the value or this portfolio will be

$DuS + rB$ with probability q,

$DS+B$

$DdS + rB$ with probability 1 - q.

Since we can select Dand B in any way we wish, suppose we choose them to equate the end-or-period values of the portfolio and the call for each possible outcome. This requires that

$$DuS + rB = C_u$$

$$DdS + rB = C_d$$

Solving these equations. we find

$$D=\frac{C_u - C_d}{(u-d)S} \qquad B = \frac{uC_d - dC_u}{(u-d)r} \qquad (1)$$

With Δ and B chosen in this way, we will call this the hedging portfolio.

If there are to be no risk less arbitrage opportunities, the current value of the call, C. cannot be less than the current value of the hedging portfolio, Δ S + *B*. If it were, we could make a riskless profit with no net investment by buying the call and selling the portfolio. It is tempting to say that it also cannot be worth more. since then we would have a risk less arbitrage opportunity by reversing our procedure and selling the call and buying the portfolio. But this overlooks the fact that the person who bought the call we sold has the right to exercise it immediately.

"Buying bonds is the same as lending; selling them is the same' as borrowing.

Suppose that Δ *S+B<S-K*. If we try to make an arbitrage profit by selling calls for more than Δ *S + B*, but less than S - *K*, then we will soon find that we are the source of arbitrage profits rather than their recipient. Anyone could make an arbitrage profit by buying our calls and exercising them immediately.

We might hope that we will be spared this embarrassment because everyone will somehow find it advantageous to hold the calls for one more period as an investment rather than take a quick profit by exercising them immediately. But each person will reason in the following way. If I do not exercise now, I will receive the same payoff as a portfolio with Δ S in stock and *B* in bonds. If I do exercise now, I can take the proceeds, S -*K*, buy this same portfolio and some extra bonds as well, and have a higher payoff in every possible circumstance. Consequently, no one would be willing to hold the calls for one more period.

Summing up all of this, we conclude that if there are to be no riskless arbitrage opportunities, it must be true that

$$C = DS + B$$

$$= \frac{C_u - C_d}{u-d} + \frac{C_d - dC_u}{(u-d)r}$$

$$= \left[\frac{(r-d)}{(u-d)} C_u + \frac{(u-r)}{(u-d)} C_d \right] \Big/ r, \qquad (2)$$

if this value is greater than S - *K*, and if not, C = S - *K*.[7]

Eq. (2) can be simplified by defining

$$p \equiv \frac{r-d}{u-d} \quad and \quad 1-p \equiv \frac{u-r}{u-d}$$

So that we can write

$$C = [pCu + (1-p)Cd]/r. \qquad (3)$$

It is easy to see that in the present case, with no dividends, this will always be greater than *S – K* as long as the interest rate is positive. To avoid

[7]1n some applications of the theory to other areas, it is useful to consider options which can be exercised only on the expiration date. These are usually termed European options. Those which can be exercised at any earlier time as well, such as we have been examining here, are then referred to as American options. Our discussion could be easily modified to include European calls. Since immediate exercise is then precluded, their values would always be given by (2), even if this is less than S -*K*.

spending time on the unimportant situations where the interest rate is less than or equal to zero, we will now assume that r is always greater than one. Hence, (3) is the exact formula for the value of a call one period prior to expiration in terms of S, *K*, u, *d*, and *r*.

To confirm this note that if *uS* • *K*, then *S* < *K* and C = 0, so *C* > *S – K*. Also if *dS* • *K*, then *C=S - (K/r) > S – K*. The remaining possibility is *uS>K>dS*. In this case, *C = p(uS – K) / r*. This is greater than *S – K* if (1 – p) *dS > (p-r) K*, which is certainly true as long as *r>1*.

This formula has a number of notable features. First, the probability *q* does not appear in the formula. This means, surprisingly, that even if different investors have different subjective probabilities about an upward or downward movement in the stock, they could still agree on the relationship of C to S. u. *d*, and *r*.

Second, the value of the call does not depend on investors' attitudes toward risk. In constructing the formula, the only assumption we made about an individual's behavior was that he prefers more wealth to less wealth and therefore has an incentive to take advantage of profitable riskless arbitrage opportunities. We would obtain the same formula whether investors are risk-averse or risk-preferring.

Third, the only random variable on which the call value depends is the stock price itself. In particular, it does not depend on the random prices of other securities or portfolios, such as the market portfolio containing all securities in the economy. If another pricing formula involving other variables was submitted as giving equilibrium market prices, we could immediately show that it was incorrect by using our formula to make risk less arbitrage profits while trading at those prices.

It is easier to understand these features .if it is' remembered that the formula is only a relative pricing relationship giving C in terms of S, U, *d*, and *r*. Investors' attitudes toward risk and the characteristics of other assets may indeed influence call values indirectly, through their effect on these variables, but they will not" be separate determinants of call value.

Finally, observe that p(*r – d) / (u – d)* is always greater than zero and less than one, so it has the properties of a probability. In fact, *p* is the value *q* would have in equilibrium if investors were risk-neutral. To see this, note that the expected rate of return on the stock would then be the riskless interest rate, so

$$q(uS) + (1-q)(dS) = rS,$$

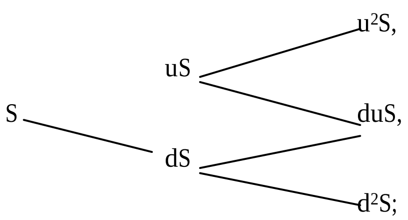
and

$$q = (r-d)/(u-d) = p.$$

Hence, the value of the call can be interpreted as the expectation of its discounted future value in a risk-neutral world. In light of our earlier observations, this is not surprising. Since the formula does not 'involve *q* or any measure of attitudes toward risk, then it must be the same for any set of, preferences, including risk neutrality.

It is important to note that this does not imply that the equilibrium expected rate of return on the call is the risk less interest rate. Indeed, our argument has shown that, in equilibrium, holding the call over the period is exactly equivalent to holding the hedging portfolio. Consequently, the risk and expected rate of return of the call must be the same as that of the hedging portfolio. It can be shown that Δ • 0 and *B* • 0, so the hedging portfolio is equivalent to a particular levered long position in the stock. In equilibrium, the same is true for the call. Of course, if the call is currently mispriced, its risk and expected return over the period will differ from that of the hedging portfolio.

Now we can consider the next simplest situation: a call with two periods remaining before its expiration date. In keeping with the binomial process, the stock can take on three possible values after two periods,



similarly, for the call,

$C_{uu} = \max [0, u^2 S - K]$

C_u

C $\qquad$ $C_{du} = \max [0, duS - K],$

C_d

$C_{dd} = \max [0, d^2 S - K]$

C_{uu} stands for the value of a call two periods from the -current time if the stock price moves upward each period; C_{du} and C_{dd} have analogous definitions.

At the end of the current period there will be one period left in the life of the call and we will be faced with a problem identical to the one we just solved. Thus, from our previous analysis, we know that when there are two periods left,

$C_u=[pC_{uu}+(1-p)C_{ud}] / r,$

and

$C_d = [pC_{du} + (1 - P)C_{dd}] / r.$

Again we can select a portfolio with Δ S in stock and *B* in bonds whose end-of-period value will be C. if the stock price goes to *uS* and C_d if the stock price goes to *d*S. Indeed, the functional form of Δ and *B* remains unchanged. To get the new values of D and *B*, we simply use eq.(1) with the new values of C_u and C_d,

Can we now say, as before, that an opportunity for profitable risk less arbitrage will be available if the current price of the call is not equal to the new value of this portfolio or S – *K*, whichever is greater? Yes, but there is an important difference. With one period to go, we could plan to lock in a risk less profit by selling an overpriced call and using part of the proceeds to buy the hedging portfolio. At the end of the period, we knew that the market price of the call must be equal to the value of the portfolio, so the entire position could be safely liquidated at that point. But this was true only because the end of the period was the expiration date. Now we have no such guarantee. At the end of the current period, when there is still one period left, the market price of the call could still be in disequilibrium and be greater than the value of the hedging portfolio. If we closed out the position then, selling the portfolio and repurchasing the call, we could suffer a loss which would more than offset our original profit. However, we could always avoid this loss by maintaining the portfolio for one more period. The value of the portfolio at the end of the current period will always be exactly sufficient to purchase the portfolio we would want to hold over the last period. In effect, we would have to readjust the proportions in the hedging portfolio, but we would not have to put up any more money.

Consequently, we conclude that even with two periods to go, there is a strategy we could follow which would guarantee risk less profits with no net investment if the current market price of a call differs from the maximum of Δ S + *Band* S – *K*. Hence, the larger of these is the current value of the call.

Since Δ and *B* have the same functional form in each period, the current value of the call in terms of C_u and C_d will again be C = $[pC_u + (1 - p)\, C_d] / r$ if this is greater than S - *K*, and C = S - *K* otherwise. By substituting from eq. (4) into the former expression, and noting that $C_{du} = C_{ud}$, we obtain

$$C=[p^2C_{uu}+2p(1-P)C_{ud}+(1-p)2C_{dd}] / r^2$$
$$=(p^2 \max[0,u^2S - K] + 2p(1-p) \max[0,duS - K]$$
$$+ (1-p)^2 \max[0, d^2 S - K] / r^2. \qquad (5)$$

A little algebra shows that this is always greater than S - *K* if, as assumed. r is always greater than one, so this expression gives the exact value of the call. [8]

All of the observations made about formula (3) also apply to formula (5), except that the number of periods remaining until expiration, *n*, now emerges clearly as an additional determinant of the call value. For formula (5), n = 2. That is, the full list of variables determining C is S, *K*, *n*, *u*, *d*, and *r*.

We now have a recursive procedure for finding the value of a call with any number of periods to go. By starting at the expiration date and working backwards, we can write down the general valuation formula for any *n:*

$$C = \left[\sum_{j=0}^{n} \left\{\frac{n!}{j!\,(n-j)!}\right\} p^j (1-p)^{n-j} \max[0, u^j d^{n-j} S - K]\right] \Big/ r^n \qquad (6)$$

This gives us the complete formula, but with a little additional effort we can express it in a more convenient way.

Let *a* stand for the minimum number of upward moves which the stock must make over the next *n* periods for the call to finish in-the-money. Thus *a* will be the smallest non-negative integer such that $u^a d^{n-a} S > K$. By taking the natural logarithm of both sides of this inequality, we could write *a* as the smallest non-negative integer greater than $1og(K/Sd^n) / log(u/d)$.

For all $j < a$,

$\max[0, u^j d^{n-j} S - K] = 0$,

and for all $j \bullet a$,

$\max[0, u^j d^{n-j} S - K] = u^j d^{n-j} S - K$.

Therefore,

$$C = \left(\sum_{j=a}^{n} \left\{ \frac{n!}{j!\,(n-j)!} \right\} p^j (1-p)^{n-j} \max[0, u^j d^{n-j} S - K] \right) \Big/ r^n$$

[8] In the current situation, with no dividends, we can show by a simple direct argument that if there are no arbitrage opportunities, then the call value must always be greater than S – *K* before the expiration date. Suppose that the call is selling for S -*K.* Then there would be an easy arbitrage strategy which would require no initial investment and would always have a positive return. All we would have to do is buy the call, short the stock, and invest *K* dollars in bonds. See Merton (1973). In the general case, with dividends, such an argument is no longer valid. and we must use the procedure of checking every period.

Of course, if $a > n$, the call will finish out-of-the-money even if the stock moves upward every period, so its current value must be zero.

By breaking up C into two terms, we can write

$$C = \left(\sum_{j=a}^{n} \left\{ \frac{n!}{j!\,(n-j)!} \right\} p^j (1-p)^{n-j} u^j d^{n-j} \left\{ \frac{u^j d^{n-1}}{r^n} \right\} \right)$$

$$- K r^{-n} \left(\sum_{j=a}^{n} \frac{n!}{j!\,(n-j)!} p^j (1-p)^{n-j} \right)$$

Now, the latter bracketed expression is the complementary binomial distri-bution function $\Phi[a; n, p]$. The first bracketed expression can also be interpreted as a complementary binomial distribution function $CP[a; n, p']$, where

$p' \equiv (u/r)p$ and $1 - p' \equiv (d/r)(1-p)$.

p' is a probability, since $0 < p' < 1$. To see this, note that $p < (r/u)$ and

$$p^j (1-p)^{n-j} \left\{ \frac{(u^j d^{n-j})}{r^n} \right\} = \frac{u}{r}^j p \, \frac{d}{r} \left((1-p) \right)^{n-j} = p'^j (1-p')^{n-1}$$

In summary:

Binomial Option Pricing Formula

$C = \Phi S[a; n, p'] - Kr^n \Phi[a; n, p]$,

where

$p \equiv (r-d)/(u-d)$ and $p' \equiv (u/r)p$,

$a \equiv$ the smallest non-negative integer greater than $\log(K/Sd^n)/\log(u/d)$

If $a > n$, $C = 0$,

It is now clear that all of the comments we made about the one period evaluation formula are valid for any number of periods. In particular, the value of a call should be the expectation, in a risk-neutral world, of the discounted value of the payoff it will receive. In fact, that is exactly what eq. (6) says, Why, then, should we waste time with the recursive procedure when we can write down the answer in one direct step? The reason is that while: this one-step approach is always technically correct, it is really useful only if we know in advance the circumstances in which a rational individual would prefer to exercise the call before the expiration date. If we do not know this, we have no way to compute the required expectation. In the present example, a call on a stock paying no dividends, it happens that we can determine this information from other sources: the call should never be exercised before the expiration date. As we will see in section 6, with puts or with calls on stocks which pay dividends, we will not be so lucky. Finding the optimal exercise strategy will be an integral part of the valuation problem. The full recursive procedure will then be necessary.

For some readers, an alternative 'complete markets' interpretation of our binomial approach may be instructive. Suppose that p_u and p_d represent the state-contingent discount rates to states *u* and *d*, respectively. Therefore, p_u would be the current price of one dollar received at the end of the period, if and only if state *u* occurs. Each security - a riskless bond, the stock, and the option -must all have returns discounted to the present by p_u and p_d if no risk less arbitrage opportunities are available. Therefore,

$$1 = \pi_u r + \pi_d r,$$

$$S = \pi_u, (uS) + \pi d\,(dS),$$

$$C = \pi_u, C_u + \pi_d C_d$$

The first two equations, for the bond and the stock, imply

$$\pi = \frac{r-d}{u-d}\,\frac{1}{r} \quad \text{and} \quad \pi_d = \frac{u-r}{u-d}\,\frac{1}{r}$$

Substituting these equalities for the state-contingent prices in the last equation for the option yields eq. (3).

It is important to realize that we are not assuming that the riskless bond and the stock and the option are the only three securities in the economy, or that other securities must follow a binomial process. Rather, however these securities are priced in relation to others in equilibrium, among themselves they must conform to the above relationships.

From either the hedging or complete markets approaches, it should be clear that three-state or trinomial stock price movements will not led to an option pricing formula based solely on arbitrage considerations. Suppose, for example, that over each period the stock price could move to *uS* or *dS* or remain the same at S. A choice of Δ and *B* which would equate the returns in two states could not in the third. That is, a risk less arbitrage position could not be taken. Under the complete markets interpretation, with three equations in now three unknown state-contingent prices, we would lack the redundant equation necessary to price one security in terms of the other two.

4. Riskless Trading Strategies

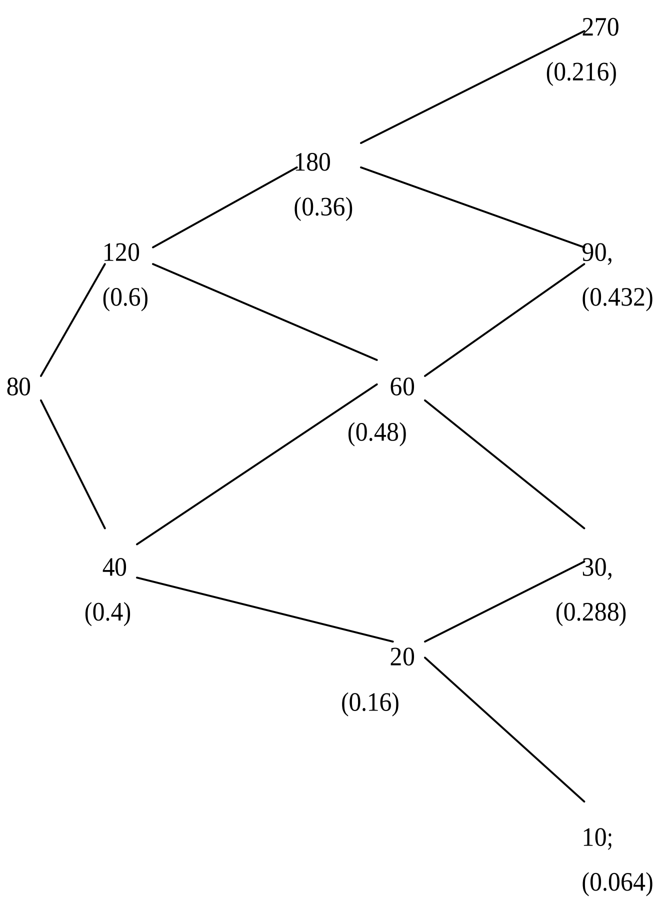
The following numerical example illustrates how we could use the formula if the current *market price* M ever diverged from its *formula value* C. If M > C, we would hedge, and if M < C, 'reverse hedge', to try and lock in a profit. Suppose the values of the underlying variables are

5=80, n = 3, K =80, u= 1.5, d =0.5, r= 1.1.

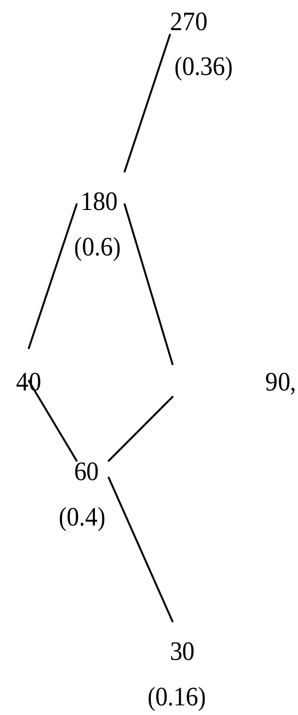
In this case, $p= (r - d)/(u - d)=0.6$. The relevant values of the discount factor are

$r^{-1}=0.909$, $r^{-2}=0.826$, $r^{-3}=0.751$.

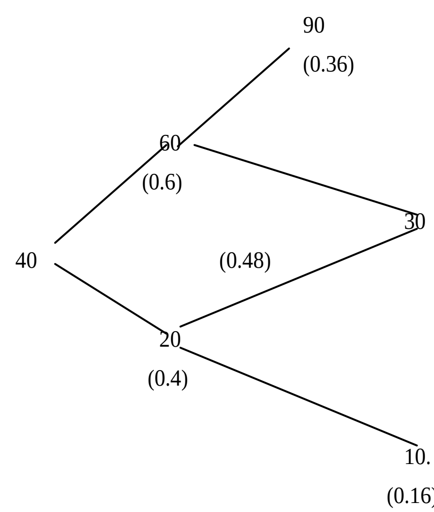
The paths the stock price may follow and their corresponding probabilities (using probability *p)* are, when n = 3, with S = 80,



when n = 2, if S = 120,



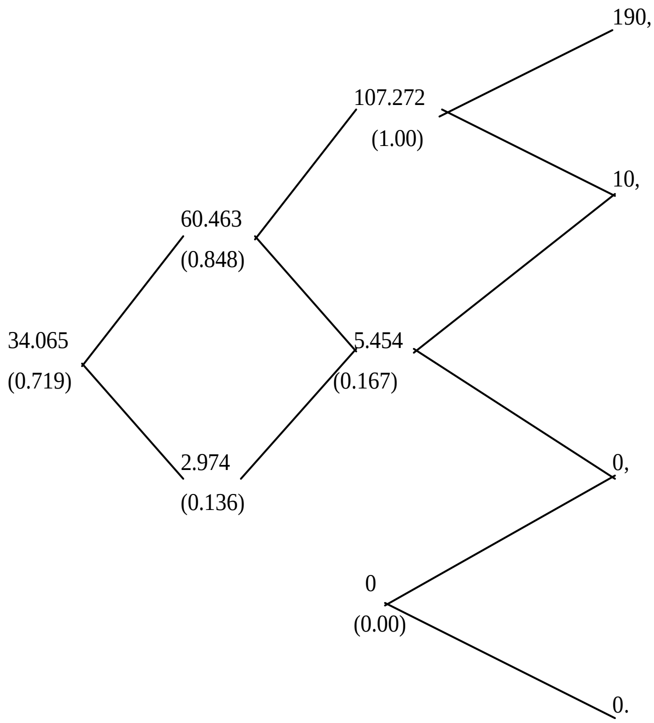
when $n=2$, if S=40,



Using the formula, the current. value of the call would be

C =0.751 [0.064(0) +0.288 (0) + 0.432(90- 80) +0.216(270 - 80)] = 34.065.

Recall that to form a riskless hedge, for each call we sell, we buy and subsequently keep adjusted a portfolio with DSin stock and *B* in bonds, where $\Delta = (C_u - C_d) / (u - d)S$. The following tree diagram gives the paths the call value may follow and the corresponding values of Δ:



With this preliminary analysis, we are prepared to use the formula to take advantage of mispricing in the market. Suppose that when n = 3, the market price of the call is 36. Our formula tells us the call should be worth 34.065. The option is over-priced, so we could plan to sell it and assure ourselves of a profit equal to the mispricing differential. Here are the steps you could take for a typical path the stock might follow.

Step 1 *(n* = 3): Sell the call for 36. Take 34.065 of this and invest it in a portfolio containing Δ =0.719 shares of stock by borrowing 0.719(80) - 34.065 = 23.455. Take the remainder, 36 - 34.065 = 1.935, and put it in the bank.

Step 2 *(n=2):* Suppose the stock goes to 120 so that the new Δ is 0.848. Buy 0.848 - 0.719 = 0.129 more shares of stock at 120 per share for a total expenditure of 15.480. Borrow to pay the bill. With an interest rate of 0.1, you already owe 23.455(1.1) = 25.801. Thus, your total current indebtedness is 25.801 + 15.480 = 41.281.

Step 3 *(n* = 1):" Suppose the stock price low goes to 60. The new t1 is 0.167. Sell 0.848-0.167=0.681 shares at 60 per share, taking in 0.681(60) = 40.860. Use this to pay back part of your borrowing. Since you now owe 41.281(1.1) = 45.409, the repayment will reduce this to 45.409-40.860=4.549.

Step 4d (n=0): Suppose the stock price now goes to 30. The call you sold has expired worthless. You own 0.167 shares of stock selling at 30 per share, for a total value of 0.167 (30) = 5. Sell the stock and repay the 4.549 (1.1) = 5 that you now owe on the borrowing. Go back to the bank and withdraw your original deposit, which has now grown to 1.935 $(1.1)^3$ = 2.575.

Step 4u (n=0): Suppose, instead, the stock price goes to 90. The call you sold is in the money at the expiration date. Buy back the call, or buy one share of stock and let it be exercised, incurring a loss of 90 - 80 = 10 either way. Borrow to cover this, bringing your current indebtedness to 5 +10 = 15. You own 0.167 shares of stock selling at 90 per share, for a total value of 0.167 (90) = 15. Sell the stock and repay the borrowing. Go back to the bank and withdraw your original deposit, which has now grown- to 1.935 $(1.1)^3$ = 2.575.

In summary, if we were correct in our original analysis about stock price movements (which did not involve the unenviable task of predicting whether the stock price would go up or down), and if we faithfully adjust our portfolio as prescribed by the formula, then we can be assured of walking away in the clear at the expiration date, while still keeping the original differential and the interest it has accumulated. It is true that closing out the position before the expiration date, which involves buying back the option at its then current market price, might produce a loss which would more than offset our profit, but this loss could always be avoided by waiting until the expiration date. More-over, if the market price comes into line with the formula value before the expiration date, we can close out the position then with no loss and be rid of the concern of keeping the portfolio adjusted.

It still might seem that we are depending on rational behavior by the person who bought the call we sold. If instead he behaves foolishly and exercises at the wrong time, could he make things worse for us as well as for himself? Fortunately, the answer is no mistakes on his part can only mean greater profits for us. Suppose that he exercises too soon. In that circum-stance, the hedging portfolio will always be worth more than S - *K,* so we could close out the position then with an extra profit.

Suppose, instead, that he fails to exercise when it would be optimal to do so. Again there is no problem. Since exercise is now optimal, our hedging portfolio will be worth S -*K.*[9] If he had exercised, this would be exactly sufficient to meet the obligation and close out the position. Since he did not, the call will be held at least one more period, so we calculate the new values of C_u and C_d and revise our hedging portfolio accordingly. But now the amount required for the portfolio, D*S* + *B,* is less than the amount we have available, S -*K.* We can withdraw these extra profits now and still maintain the hedging portfolio. The longer the holder of the calI goes on making mistakes, the better off we will be.

[9]If we were reverse hedging by buying an undervalued call and selling the hedging portfolio, then we would ourselves want to exercise at this point. Since we will receive S - *K* from exercising, this will be exactly enough money to buy back the hedging portfolio.

Consequently, we can be confident that things will eventually work out right no matter what the other party does. The return on our total position, when evaluated at prevailing market prices at intermediate times, may be negative. But over a period ending no later than the expiration date, it win be positive.

In conducting the hedging operation, the essential thing was to maintain the proper proportional relationship: for each call we are short, we hold J shares of stock and the dollar amount *B* in bonds in the hedging portfolio. To emphasize this, we will refer to the number of shares held for each call as the hedge ratio. In our example, we kept the number of calls constant and made adjustments by buying or selling stock and bonds. As a result, our profit was independent of the market price of the call between the time we initiated the hedge and the expiration date. If things got worse before they got better, it did not matter to us.

Instead, we could have made the adjustments by keeping the number of shares of stock constant and buying or selling calls and bonds: However, this could be dangerous. Suppose that after initiating the position, we needed to increase the hedge ratio to maintain the proper proportions. This can be achieved in two ways:

a. buy more stock, or

b. buy back some of the calls.

If we adjust through the stock, there is no problem. If we insist on adjusting through the calls, not only is the hedge no longer risk less, but it could even end up losing money! This can happen if the call has become even more overpriced. We would then be closing out part of our position in calls at a loss. To remain hedged, the number of calls we would need to buy back depends on their value, not their price. Therefore, since we are uncertain about their price, we then become uncertain about the return from the hedge.

Worse yes, if the call price gets high enough, the loss on the closed portion of our position could throw the hedge operation into an overall loss.

To see how this could happen, let us rerun the hedging operation, where we adjust the hedge ratio by buying and selling calls.

Step 1 (n = 3): Same as before.

Step 2 (n =2): Suppose the stock goes to 120, so that the new J =0.848. The call price has gotten further out of line and is now selling for 75. Since its value is 60.463, it is now overpriced by 14.537. With 0.719 shares you must buy back 1-0.848=0.152 calls to produce a hedge ratio of 0.848 =0.719/0.848. This costs 75(0.152) =11.40. Borrow to pay the bill. With the interest rate of 0.1, you already owe 23.455(1.1) = 25.801. Thus, your total current indebtedness is 25.801 + 11.40 = 37.201.

Step 3 (n = 1): Suppose the stock goes to 60 and the call is selling for 5.454. Since the call is now fairly valued, no further excess profits can be made by continuing to hold the position. Therefore, liquidate by selling your 0.719 shares for 0.719(60) = 43.14 and close out the call position by buying back 0.848 calls for 0.848(5.454) = 4.625. This nets 43.14-4.625=38.515. Use this to pay back part of your borrowing. Since you now owe 37.200.1) = 40.921, after repayment you owe 2.406. Go back to the bank and withdraw your original deposit, which has now grown to $1.935(1.1)^2$ = 2.341. Unfortunately, after using this to repay your remaining borrowing, you still owe 0.065.

Since we adjusted our position at Step 2 by buying overpriced calls, our profit is reduced. Indeed, since the calls were considerably overpriced, we actually lost money despite apparent profitability of the position at Step 1. We can draw the following adjustment rule from our experiment: *To adjust a hedged position, never buy an overpriced option or sell an under priced option.* As a corollary, whenever we can adjust a hedged position by buying more of an underpriced option or selling more of an overpriced option, our profit will be enhanced if we do so. For example, at Step 3 in the original hedging illustration, had the call still been overpriced, it would have been better to adjust the position by selling more calls rather than selling stock. In summary, by choosing the right side of the position to adjust at intermediate dates, *at a minimum* we can be assured of earning the original differential and its accumulated interest, and we may earn considerably more.

5. Limiting Cases

In reading the previous sections, there is a natural tendency to associate with each period some particular length of calendar time, perhaps a day. With this in mind, you may have had two objections. In the first place, prices a day from now may take on many more than just two possible values. Furthermore, the market is not open for trading only once a day, but, instead, trading takes place almost continuously.

These objections are certainly valid. Fortunately, our option pricing approach has the flexibility to meet them. Although it might have been natural to think of a period as one day, there was nothing that forced us to do so. We could have taken it to be a much shorter interval – say an hour -or even a minute. By doing so, we have met both objections simultaneously. Trading would take place far more frequently, and the stock price could take on hundreds of values by the end of the day.

However, if we do this, we have to make some other adjustments to keep the probability small that the stock price will change by a large amount over a minute. We do not want the stock to have the same percentage up and down moves for one minute as it did before for one day. But again there is no need for us to have to use the same values. .We could, for example, think of the price as making only a very small percentage change over each minute.

To make this more precise, suppose that h represents the elapsed time between successive stock price changes. That is, if t is the fixed length of calendar time to expiration, and n is the number of periods of length h prior to expiration, then

$$h \equiv t/n.$$

As trading takes place more and more frequently, h gets closer and closer to zero. We .must then adjust the interval-dependent variables r, u, and d in such a way that. we obtain empirically realistic results as h becomes smaller, or, equivalently, as $n \rightarrow \infty$

When we were thinking of the periods as having a fixed length, r represented both the interest rate over a fixed length of calendar time and the interest rate over one period. Now we need to make a distinction between these two meanings. We will let r continue to mean one plus the interest rate over a fixed length of calendar time. When we have occasion to refer to one plus the interest rate over a period (trading interval) of length h, we will use the symbol .

Clearly, the size of depends on the number of subintervals, n, into which t is divided. Over the n periods until expiration, the total return is ^{n}, where n = t/h. Now not only do we want to depend on *n*, but we want it to depend on 11 in a particular way - so that as *n* changes the total return over the fixed time *t.* remains the same. This is because the interest rate obtainable over some fixed length of calendar time should have nothing to do with how we choose to think of the length of the time interval *h.*

If r (without the 'hat') denotes one plus the rate of interest over a *fixed* unit, of calendar time, then over elapsed time *t,* r' is the total return.[10] Observe that this measure of total return does not depend on *n.* As we have argued, we want to choose the dependence of on *n,* so that

$^n = r^t$,

for any choice of *n*. Therefore, $= r^{t/n}$ This last equation shows how must depend on *n* for the total return over elapsed time *t* to be independent of *n.*

We also need to define *u* and *d* in terms of *n.* At this point, there are two significantly different paths we can take. Depending on the definitions we choose, as $n \to \infty$ (or, equivalently, as $h \to 0$), we can have either a continuous or a jump stochastic process. In the first situation very small random changes in the stock price will be occurring in each very small time interva1. The stock price will fluctuate incessantly, but its path can be drawn without lifting pen from paper. In contrast, in the second case, the stock price will usually move in a smooth deterministic way, but will occasionally experience sudden discontinuous changes. Both can be derived from our binomial process simply by choosing how *u* and *d* depend on *n.* We examine in detail only the continuous process which leads 'to the option pricing formula originally derived by Fischer Black and Myron Scholes, Subsequently, we indicate how to develop the jump process formula originally derived by John Cox and Stephen Ross.

[10]The. scale of this unit (perhaps a day, or a year) is unimportant as long as r and *I* are expressed in the same scale.

Recall that we supposed that over each period the stock price would experience a one plus rate of return of *u* with probability *q* and *d* with probability 1 - *q.* It will be easier and clearer to work, instead, with the natural logarithm of the one plus rate of return, log *u* or log *d.* This gives the continuously compounded rate of return on the stock over each period. It is a random variable which, in each period, will be equal to log *u* with probability *q* and log *d* with probability 1 - *q.*

Consider a typical sequence of five moves, say *u, d, u, u, d.* Then the final stock price will be $S^* = uduudS$; $S^*/S = u^3 d^2$, and $\log(S^*/S) = 3\log u + 2\log d$, More generally, over *n* periods,

$$\log(S^*/S) = j\log u + (n - j)\log d = j\log(u/d) + n\log d,$$

where j is the (random) number of upward moves occurring during the 11 periods to expiration. Therefore, the expected value of $\log(S^*/S)$ is

$$E[\log(S^*/S)] = \log(u/d)\; E(j) + n\log d,$$

and its variance is

$$\operatorname{var}[\log(S^*/S)] = [\log(u/d)]^2 \cdot \operatorname{var}(j).$$

Each of the *n* possible upward moves has probability *q.* Thus, $E(j) = nq$. Also, since the variance each period is $q(1-q)^2 + (1-q)(O-q)^2 = q(1-q)$, then $\operatorname{var}(j) = nq(1-q)$. Combining all of this, we have

$$E[\log(S^*/S)] = [q\log(u/d) + \log d]\, n = \mu n,$$

$$\operatorname{var}[\log(S^*/S)] = q(1-q)[\log(u/d)]^2 n \equiv \sigma n.$$

Let us go back to our discussion. We were considering dividing up our original longer time period (a day) into many shorter periods (a minute or even less). Our procedure calls for, over fixed length of calendar time *t* making *n* larger and larger. Now if we held everything else constant while we let *n* become large, we would be faced with the problem we talked about earlier. In fact, we would certainly not reach a reasonable conclusion if either μn or s went to zero or infinity as *n* became large. Since *t* is a fixed length of time, in searching for a realistic result, we must make the appropriate adjustments in *u. d,* and *q.* In doing that, we would at least want the mean and variance of the continuously compounded rate of return of the assumed stock price movement to coincide with that of the actual stock price as $n \to \infty$, Suppose we label the actual empirical values of μn and $s^2 n$ as μt and $s^2 t$, respectively. Then we would want to choose *u, d,* and *q,* so that

$$[q\log(u/d) + \log d]n \to \mu t$$

$$\text{as } n \to \infty$$

$$q(1-q)[\log(u/d)]^2 n \to s$$

A little algebra shows we can accomplish this by letting

$$u = e^{\sigma\sqrt{t/n}}, \quad d = e^{-\sigma\sqrt{t/n}}, \quad q = \tfrac{1}{2} + \tfrac{1}{2}(\mu/\sigma)\sqrt{t/n}$$

In this case, for any n

$$\mu n = \mu t \;\text{ and }\; s^2 n = [s^2 - \mu^2 (t/n)]t$$

Clearly, as $n \to \infty, \hat{s}^2 n$, while $\hat{\mu} n = \mu t$ for all values of n

Alternatively, we could have chosen *u, d,* and *q* so that the mean and variance of the future stock price for the discrete binomial process approach the pre specified mean and variance of the actual stock price as $n \to \infty$. However, just as we would expect, the same values will accomplish this as well. Since this would not change our conclusions, and it is computationally more convenient to work with the continuously compounded rates of return, we will proceed in that way.

This satisfies our initial requirement that the limiting means and variances coincide, but we still need to verify that we are arriving at a sensible limiting probability distribution of the continuously compounded rate of return. The mean and variance only describe certain aspects of that distribution. For our model, the random" continuously compounded rate of return over a period of length *t* is the sum of *n* independent random variables, each of which can take the value log *u* with probability *q* and log *d* with probability 1 - *q.* We wish to know about the distribution of this sum as *n* becomes large and *q, u,* and *d* are chosen in the way described. We need to remember that as we change *n,* we are not simply adding one more random variable to the previous sum, but instead are changing the probabilities and possible outcomes for every member of the sum. At this point, we can rely on a form of the central limit theorem which, when applied to our problem, says that, as $n \to \infty$, if

$$\frac{[\log u - \mu]^3 + (1-q)[\log d - \mu]^3}{\hat{\sigma}^3 \sqrt{n}} \rightarrow 0$$

then

$$\text{Prob}\left(\left\{\frac{\log(S^*/S) - \hat{\mu} n}{\sigma\sqrt{n}}\right\} \Box z\right) \rightarrow N(z)$$

where $N(z)$ is the standard normal distribution functio'1. Putting this into words, as the number of periods into which the fixed length of time to expiration is divided approaches infinity, the probability that the standar-dized continuously compounded rate of return of the stock through the expiration date is not greater than the number z approaches the probability under a standard normal distribution.

The initial condition says roughly that higher-order properties of the distribution, such as how it is skewed, become less and less important, relative to its standard deviation, as $n \rightarrow \infty$. We can verify that the condition is satisfied by making the appropri-ate substitutions and finding

$$\frac{q[\log u - \hat{\mu}]^3 + (1-q)[\log d - m]^3 - (1-\hat{q})^2 + q^2}{\hat{s}^3 \ddot{O} n}$$

then

$$\text{Prob}\left(\left\{\frac{\log(S^*/S) - \mu n}{\sigma\sqrt{n}}\right\} \bullet z \rightarrow N(z),\right)$$

where N (z) is the standard normal distribution function putting this into words as the number of periods into which the fixed length of time to expiration is divided approaches infinity, the probability that the standardized continuously compounded rate of return of the stock through the expiration date is not greater than the number z approaches the probability under a standard normal distribution.

The initial condition says roughly that higher-order properties of the distribution, such as how it is skewed, become less and less important, relative to its standard deviation, as $n \rightarrow \infty$. We can verify that the condition is satisfied by making the appropri-ate substitutions and finding

$$\frac{q[\log u - \mu]^3 + (1-q)\ [\log d - \mu]^3}{\hat{\sigma}^3\sqrt{n}} \quad \frac{(1-q)^2 + q^2}{\sqrt{nq(1-q)}}$$

which goes to zero as $n \rightarrow \infty$ since $q = ½ + ½ (\mu/s) \sqrt{\ } \ t/n$

Thus, the multiplicative binomial model for stock prices includes the lognormal distribution as a limiting case.

Black and' Scholes began directly with continuous trading and the assumption of a lognormal distribution for stock prices. Their approach relied on some quite advanced mathematics. However, since our approach contains continuous trading and the lognormal distribution as a limiting case, the two resulting formulas should then coincide. We will see shortly that this is indeed true, and we will have the advantage of using a much simpler method. It is important to remember, however, that the economic arguments we used to link the option value and the stock price are exactly the same as those advanced by Black and Scholes (1973) and Merton (1973, 1977.

The formula derived by Black and Scholes, rewritten in terms of our notation, is

Black -Scholes Option Pricing Formula

$$C = SN(x) - Kr^{-1} N(x - s\ddot{O}t),$$

where

$$\frac{\log(S/Kr^{-t})}{s\ddot{O}t} + ½\, s\ddot{O}t$$

We now wish to confirm that our binomial formula converges to the Black-Scholes formula when t is divided into more and more subintervals, and ,u, d. and q are chosen in the way we described that is, in a way such that the multiplicative binomial probability distribution of stock prices goes to the lognormal distribution.

For easy reference, let us recall our binomial option pricing formula:

$$C = Sf[a; n, p'] - K \bullet^{-n} \ f[a; n, p].$$

The similarities are readily apparent. $^{-n}$ is, of course, always equal to r^{1}. Therefore, to show the two formulas conve_ge, we need only show that as $n \rightarrow \infty$.

$f[a; n, p'] \rightarrow N(x)$ and $\phi[a; n, p] \rightarrow N(x - s\ddot{O}t)$.

We will consider only $f[a;n,p]$, since the argument is exactly the same for $f[a; n, p']$.

The complementary binomial distribution function $f[a;n,p]$ is the prob-ability that the sum of n random variables, each of which can take on the value I with probability p and 0 with probability 1 - p, will be greater than or equal to a. We know that the random value of this sum, j, has mean np and standard deviation $\ddot{O}np(1-p)$. Therefore,

$$1 - \phi[a; n, p] = \text{Prob}[\, j \bullet a - 1]$$

$$= \text{Prob}\ \frac{j - np}{\ddot{O}np(1-p)} \quad \frac{a - 1 - np}{\ddot{O}np(1-p)}$$

Now we can make an analogy with our earlier discussion. If we consider a stock which in each period will move to uS with probability p and dS with probability 1 – p, then $\log(S^*/S) = j \log(u/d) + n \log d$. The mean and variance of the continu-ously compounded rate of return of this stock are

$m_p = p \log(u/d) + \log d$ and $\sigma^2_p = p(1-p)[\log(u/d)]^2$

Using these equalities, we find that

$$\frac{j - np}{np(1-p)} \quad \frac{\log(S^*/S) - m_p n}{\hat{s}_p \ddot{O} n}$$

Recall from the binomial formula that

$$a - I = \log(K/Sd^n)/\log(u/d) - e$$

$$= [\log(K/S) - n \log d] \log(u/d) - e$$

where e is a number between zero and one. Using this and the definitions of μ_p and σ^2_p, with a little algebra, we have

$$\frac{a - 1 - np}{np(1-p)} = \frac{\log(K/S) - \hat{\mu}_p n - e \log(u/d)}{\hat{\sigma}_p \sqrt{n}}$$

Putting these results together,

$1 - \phi\,[a;\, n,\, p]$

$$= \text{Prob}\left(\frac{\log(S^*/S) - \hat{\mu}_p n}{\hat{\sigma}_p \sqrt{n}} \geq \frac{\log(K/S) - \hat{\mu}_p n - \varepsilon \log(u/d)}{\hat{\sigma}_p \sqrt{n}} \right)$$

We are now in a position to apply the central limit theorem. First, we must check if the initial condition,

$$\frac{p[\log u - \hat{\mu}_p]^3 + (1-p)[\log d - \hat{\mu}_p]^3}{\hat{\sigma}_p \sqrt{n}} = \wedge \frac{(1-p)^2 + p^2}{\sqrt{np(1-p)}} \rightarrow 0$$

as $n \rightarrow \infty$, is satisfied. By first recalling that $p \equiv (r - d)/(u-d)$, and then $r = r^{t/n}$, $u = e^{\sigma\sqrt{t/n}}$ and $d = e^{\sigma\sqrt{t/n}}$, it is possible to show that as $n \rightarrow \infty$,

$$P \rightarrow \tfrac{1}{2} + \tfrac{1}{2}\left(\log r - \tfrac{1}{2}\sigma^2 / \sigma\right)\sqrt{t/n}$$

As a result, the initial condition holds, and we are justified in applying the central limit theorem.

To do so, we need only evaluate $\hat{\mu}_p n$, $\hat{\sigma}^2_p n$ and $\log(u/d)$ as $n \rightarrow \infty$.[11] Examination of our discussion for parameterizing q shows that as $n \rightarrow \infty$,

$$\hat{\mu}_p n \rightarrow (\log r - \tfrac{1}{2}\sigma^2)t \quad \text{and} \quad \hat{\sigma}_p \sqrt{n} \rightarrow \sigma\sqrt{t}.$$

Furthermore, $\log(u/d) \rightarrow 0$ as $n \rightarrow \infty$.

For this application of the central limit theorem, then, since

$$\frac{\log(K/S) - \hat{\mu}_p n - \varepsilon \log(u/d)}{\hat{\sigma}_p \sqrt{n}} \rightarrow z = \frac{\log(K/S) - (\log r - \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}},$$

we have

$$1 - \Phi[a; n, p] \rightarrow N(z) = N\left[\frac{\log(Kr^{-t}/S)}{\sigma\sqrt{t}} + \tfrac{1}{2}\sigma\sqrt{t}\right].$$

The final step in the argument is to use the symmetry property of the standard normal distribution that $1 - N(z) = N(-z)$. Therefore, as $n \rightarrow \infty$,

$$\Phi[a; n, p] \rightarrow N(-z) = N\left[\frac{\log(S/Kr^{-t})}{\sigma\sqrt{t}} - \tfrac{1}{2}\sigma\sqrt{t}\right] = N(x - \sigma\sqrt{t}).$$

Since a similar argument holds for $\Phi[a; n, p']$, this completes our dem-

[11] A surprising feature of this evaluation is that although $p \neq q$ and thus $\hat{\mu}_p \neq \hat{\mu}$, and $\hat{\sigma}_p \neq \hat{\sigma}$, nonetheless $\hat{\sigma}_p \sqrt{n}$ and $\hat{\sigma}\sqrt{n}$ have the same limiting value as $n \rightarrow \infty$. By contrast, since $\mu \neq \log r - (\frac{1}{2}\sigma^2)$, $\hat{\mu}_p n$ and $\hat{\mu} n$ do not. This results from the way we needed to specify u and d to obtain convergence to a lognormal distribution. Rewriting this as $\sigma\sqrt{t} = (\log u)\sqrt{n}$, it is clear that the limiting value σ of the standard deviation does not depend on p or q, and hence must be the same for either. However, at any point before the limit, since

$$\hat{\sigma}^2 n = \left(\sigma^2 - \mu^2 \frac{t}{n}\right)t \quad \text{and} \quad \hat{\sigma}^2_p n = \left[\sigma^2 - (\log r - \tfrac{1}{2}\sigma^2)^2 \frac{t}{n}\right]t,$$

$\hat{\sigma}$ and $\hat{\sigma}_p$ will generally have different values.

The fact that $\hat{\mu}_p n \rightarrow (\log r - \frac{1}{2}\sigma^2)t$ can also be derived from the property of the lognormal distribution that

$$\log E[S^*/S] = \mu_p t + \tfrac{1}{2}\sigma^2 t,$$

where E and μ_p are measured with respect to probability p. Since $p \doteq (\hat{r} - d)/(u - d)$, it follows that $\hat{r} = pu + (1-p)d$. For independently distributed random variables, the expectation of a product equals the product of their expectations. Therefore,

$$E[S^*/S] = [pu + (1-p)d]^n = \hat{r}^n = r^t.$$

Substituting r^t for $E[S^*/S]$ in the previous equation, we have

$$\mu_p = \log r - \tfrac{1}{2}\sigma^2.$$

onstration that the binomial option pricing formula contains the Black–Scholes formula as a limiting case.[12,13]

As we have remarked, the seeds of both the Black–Scholes formula and a continuous-time jump process formula are both contained within the binomial formulation. At which end point we arrive depends on how we take limits. Suppose, in place of our former correspondence for u, d, and q, we instead set

$$u = u, \qquad d = e^{\zeta(t/n)}, \qquad q = \lambda(t/n).$$

This correspondence captures the essence of a pure jump process in which each successive stock price is almost always close to the previous price $(S \to dS)$, but occasionally, with low but continuing probability, significantly different $(S \to uS)$. Observe that, as $n \to \infty$, the probability of a change by d becomes larger and larger, while the probability of a change by u approaches zero.

With these specifications, the initial condition of the central limit theorem we used is no longer satisfied, and it can be shown the stock price movements converge to a log-Poisson rather than a lognormal distribution as $n \to \infty$. Let us define

$$\Psi[x; y] \equiv \sum_{i=x}^{\infty} \frac{e^{-y} y^i}{i!},$$

[12]The only difference is that, as $n \to \infty$, $p' \to \frac{1}{2} + \frac{1}{2}[(\log r + \frac{1}{2}\sigma^2)/\sigma]\sqrt{t/n}$.

Further, it can be shown that as $n \to \infty$, $\Delta \to N(x)$. Therefore, for the Black–Scholes model, $\Delta S = SN(x)$ and $B = -Kr^{-t}N(x - \sigma\sqrt{t})$.

[13]In our original development, we obtained the following equation (somewhat rewritten) relating the call prices in successive periods:

$$\left(\frac{\hat{r} - d}{u - d}\right) C_u + \left(\frac{u - \hat{r}}{u - d}\right) C_d - \hat{r} C = 0.$$

By their more difficult methods, Black and Scholes obtained directly a partial differential equation analogous to our discrete-time difference equation. Their equation is

$$\tfrac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + (\log r) S \frac{\partial C}{\partial S} - \frac{\partial C}{\partial t} - (\log r) C = 0.$$

The value of the call, C, was then derived by solving this equation subject to the boundary condition $C^* = \max[0, S^* - K]$.

Based on our previous analysis, we would now suspect that, as $n \to \infty$, our difference equation would approach the Black–Scholes partial differential equation. This can be confirmed by substituting our definitions of $\hat{r}$, u, d in terms of n in the way described earlier, expanding C_u, C_d in a Taylor series around $(e^{\sigma\sqrt{h}}S, t - h)$ and $(e^{-\sigma\sqrt{h}}S, t - h)$, respectively, and then expanding $e^{\sigma\sqrt{h}}$, $e^{-\sigma\sqrt{h}}$, and r^h in a Taylor series, substituting these in the equation and collecting terms. If we then divide by h and let $h \to 0$, all terms of higher order than h go to zero. This yields the Black–Scholes equation.

as the complementary Poisson distribution function. The limiting option pricing formula for the above specifications of *u*, *d*, and *q* is then

Jump Process Option Pricing Formula

> $C = ?\ [x: y] - Kr^{-1}\ ?\ [x; y\ /\ u],$
>
> where
>
> $y\ ?\ (\log r - ?)\ ut/(u - 1),$
>
> and
>
> x ? the smallest non-negative integer greater than (log(K/S) - ? *t* / log *u*.

A very similar formula holds if we let $u = e^{?\ (t/n)}$, $d = d$, and $1 - q = ?\ (t/n)$.

6. Dividends and Put Pricing

So far we have been assuming that the stock pays no dividends. It is easy to do away with this restriction. We will illustrate this with a specific dividend policy: the stock maintains a constant yield, *d* , on each ex-dividend date. Suppose there is one period remaining before expiration and the current stock price is S. If the end of the period is an ex-dividend date, then an individual who owned the stock during the period will receive at that time a dividend of either *d uS or d dS*. Hence, the stock price at the end of the period will be either $u\,(1 - d)$ S, or $d\,(1 - d)$ S, where v = 1 if the end of the period is an ex-dividend date and v = 0 otherwise. Both *d* and v are assumed to be known with certainty.

When the call expires, its contract and a rational exercise policy imply that its value must be either

$C_u = \max\,[0, u(1-d)^v S - K],$

or

$C_d = \max\,[0, d\,(1 - d)^v S - K].$

Therefore,

$C_u = \max\,[0, u\,(1 - d)^v S - K],$

C

$C_d = \max\,[0, d\,(1 - d)^v S - K]$

Now we can proceed exactly as before. Again we can select a portfolio of Δ shares of stock and the dollar amount B in bonds which will have the same end-of-period value as the call.[14] By retracing our previous steps, we can show that

$C = [pC_u + (1 + P)\,C_d\ /\ \ ,$

If this is greater than S – K and C=S – K otherwise. Here, once again $= (\ \ -d)\ /\ (u - d)$ and $\Delta = (C_u - C_d)\ /\ (u - d)\ /\ S$.

Thus far the only change is that $(1 - d)^v$ S in the values for C_u and C_{-d}. Now we come to the major difference: early exercise may be optimal. To see this, suppose that v = 1 and d (1- *d*) S > K. Since u > d, then, also, u (1 - *d*) S > K. In this case, C_u = u (1 - *d*) S - K and C_d = d (1 - *d*) S - K. Therefore, since (u /) p + (d/) (1 – p) = 1, $[pC_u + (1 - p)\,C_d]\ /\ \ = (1 - d)$ S - (K/). For sufficiently high stock prices, this can obviousl be less than S – K. Hence, there are definitely some circumstances in which no one would be willing to hold the call for one more period.

In fact, there will always be a critical stock price, S such that if S >, the call should be exercised immediately. Ŝ will be the stock price at which $[pC_{u+}(1 - p)C_d]\ /\ \ = S - K$.[15] That is it is the lowest stock price at which the value of the hedging portfolio exactly equals S – K. This means Ŝ will, other things equal, be lower the striking price.

We can extend the analysis to an arbitrary number of periods in the same way as before. There is only one additional difference, a minor modification in the heading operation. Now the funds in the hedging portfolio will be increased by any dividends received, or decreased by the restitution required for dividends paid while the stock is held short.

Although the possibility of optimal exercise before the expiration date causes no conceptual difficulties, it does seem to prohibit a simple closed form solution for the value of a call with many periods to go. However, our analysis suggests a sequential numerical procedure which will allow us to calculate the continuous time value to any desired degree of accuracy.

Let C be the current value of a call with n periods remaining. Define

$$v(n, i) \equiv \sum_{k=1}^{n-i} v_k$$

So that $\bar{v}\ (n - i)$ is the number of ex-dividend dates occurring during the next $n - i$ periods from now, given that the current stock price S has changed to u jd n – i - j($1 - d)^{\bar{v}(n-i)}$ S, where $j = 0, 1, 2, ..., n - i$.

With this notation, we are prepared to solve for the current value of the call by working backward in time from the expiration date. At expiration, $i = 0$, so that

$C(n, 1, j) = \max\,[0, u^j d^{n-j} - (1 - d)^{v(n, 0)}\,S - K]$ for j =0, 1,..., n.

One period before the expiration date, $i = 1$ so that

$C(n, 1, j) = \max[u^j d^{n-1-}\ (1 - d)^{v(n,1)}\,S - K,$

$[pC\,(n, 0, j + 1) + (1 - p)\,C\,(n, 0, j)]\ /\ \bullet\ ]$

for j=0, 1,...,n – 1 .

More generally, *i* periods before expiration

$C\,(n, i, j) = \max\,[u^j d^{n-1-j}\,(1 - d)^{v(n, i)}\ S - K,$

$[pC\,(n, i-1, j + 1) + (1 - p)C\,(n, i-1, j)]\ /\ \bullet\ ]$

for j = 0,1,…., n – i

Observe that each prior step provides the inputs needed to evaluate the right-hand arguments of each succeeding step. The number of calculations decreases as we move backward in time. Finally, with *n* periods before expiration, since $i = n$, C = *C(n, n,* 0)= max[S -*K, [pC(n, n* -1,1) + (1- *p)C(n,* n -1, 0)] / •],

and the hedge ratio is

$$\Delta = \frac{C(n, n-1, 1) - C(n, n-1, 0)}{(u-d)S}.$$

We could easily expand the analysis to include dividend policies in which the amount paid on any ex-dividend date depends on the stock price at that time in a more general way.[16] However; this will cause some minor complications. In our present example with a constant dividend yield, the possible stock prices n - *i* periods from now are completely determined by the total number of upward moves (and ex-dividend dates) occurring during that interval. With other types of dividend policies, the enumeration will be more complicated, since then the terminal stock price will be affected by the timing of the upward moves as well as their total number. But the basic principle remains the same. We go to the expiration date and calculate the call value for all of the possible prices that the stock could have then. Using this information, we step back one period and calculate the call values for all possible stock prices at that time, and so forth.

We will now illustrate the use of the binomial numerical procedure in approximating continuous-time call values. In order to have an exact continuous-time formula to use for comparison, we will consider the case with no dividends. Suppose that we are given the inputs required for the Black-Scholes option pricing formula: S, *K, t,* s , and *r.* To convert this information into the inputs *d, u,* and required for the binomial numerical procedure, we use the relationships:

$d = 1 / u,$ $u = e^{s\sqrt{t/n}}$ $\bullet = r^{t/n}$

Table 2 gives us a feeling for how rapidly option values approximated by the binomial method approach the corresponding limiting Black-Scholes values given by $n = \infty$. At n = 5, the values differ by at most $0.25; and at n = 20, they differ by at most $0.07. Although not shown, at n = 50, the greatest difference is less than $0.03, and at n = 150, the values are identical to the penny.

To derive a method for valuing puts, we again use the binomial for-mulation. Although it has been convenient to express the argument in terms of a particular security, a call, this is not essential in any way. The same basic analysis can be applied to puts.

Letting P denote the current price of a put, with one period remaining before expiration, we have

$$P \begin{cases} P_u = \max[0, K - u(1-d)^v S], \\ P_d = \max[O, K - d(1-d)^v S]. \end{cases}$$

Once again, we can choose a portfolio with ΔS in stock and *B* in bonds which will have the same end-of-period values as the put. By a series of steps which are formally equivalent to the ones, we followed in section 3, we can show that

P = $[pP_u + (1 - p) P_d] / \bullet$,

if this is greater than *K* - S, and P =*K* - S otherwise. As before, $p = (\ - d) / (u-d)$ and $A = (P_u - P_d) / (u-d) S$ . . Note that for puts, since $P_u \bullet P_d$, then $\Delta \bullet 0$. This means that if we sell an overvalued put, the hedging portfolio which we buy will involve a short position in the stock.

We might hope that with puts we will be spared the complications caused by optimal exercise before the expiration date. Unfortunately, this is not the case. In fact, the situation is even worse in this regard. Now there are always some possible circumstances in which no one would be willing to hold the put for one more period.

To see this, suppose $K > u(1 - d)^v S$. Since $u > d$, then, also, $K > d(1 - d)^v S$. In this case, $P_u = K - u(1 - d)^v S$ and $P_d = K - d(1 - d)^v S$. Therefore, since $(u/\ )p + (d/\ )(1 - p) = 1$,

$[pP_u + (1 - p)P_d] / \bullet = (K/\bullet) - (1 - d)^v S.$

If there are no dividends (that is, v = 0), then this is certainly less than *K* - S. Even with v = 1, it will be less for a sufficiently low stock price.

Thus, there will now be a critical stock price, $\hat{S}$, such that if S < $\hat{S}$, the put should be exercised immediately. By analogy with our discussion for the call, we can see that this is the stock price at which $[pP_u + (1 - p)P_d] / \ = K - S.$ Other things equal, S will be higher the lower the dividend yield, the higher the interest rate, and the higher the striking price. Optimal early exercise thus becomes more likely if the put is deep-in-the-money and the interest rate is high. The effect of dividends yet to be paid diminishes the advantages of immediate exercise, since the put buyer will be reluctant to sacrifice the forced declines in the stock price on future ex-dividend dates.

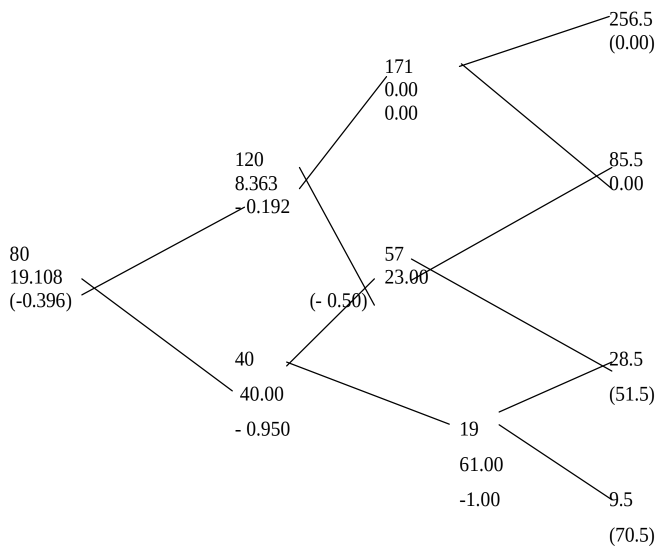
This argument can be extended in the same way as. Before to value puts with any number of periods to go. However, the chance for optimal exercise before the expiration date once again seems to preclude the possibility of expressing this value in a simple form. But our analysis also indicates that, with slight modification, we can value puts with the same numerical techniques' we use for calls. Reversing the difference between the stock price and the striking price at each stage is the only change.[17]

[17] Michael Parkinson (1977) has suggested a similar numerical procedure based on a trinomial process, where the stock price can either increase, decrease, or remain unchanged. In fact, given the theoretical basis for the binomial numerical procedure provided, the numerical method can be generalized to permit $k + 1 \bullet n$ jumps to new stock prices in each period. We can consider exercise only every k periods, using the binomial formula to leap across intermediate periods. In effect, this means permitting k + 1 possible new stock prices before exercise is again considered. That is, instead of considering exercise n times, we would only consider it about n/k times. For fixed t and k, as $n \to \infty$, option values will approach their continuous-time values.

This alternative procedure is interesting, since it may enhance computer efficiency. At one extreme, for calls on stocks which do not pay dividends setting k + I = n gives the most efficient results. However, when the effect of potential early exercise is important and greater accuracy is required. the most efficient results are achieved by setting $k = 1$ as in our description above.

The diagram presented in table 3 shows the stock prices, put values, and values of Δ obtained in this way for the example given in section 4. The values used there were S=80, K=80, n=3, u=1.5, d=0.5, and $= 1.1$. To include dividends as well, we assume that a cash dividend of five percent ($d = 0.05$) will be paid at the end of the last period before the expiration date. Thus, $(1 - d)^{v(n,0)} = 0.95$, $(1 - d)^{v(n,1)} = 0.95$, and $(1 - d)^{v(n,2)} = 1.0$. Put values in italics indicate that immediate exercise is optimal.

Table – 3



7. Conclusion

It should now be clear that whenever stock price movements conform to a discrete binomial process, or to a limiting form of such a process, options can be priced solely on the basis of arbitrage considerations. Indeed, we could have significantly complicated the simple binomial process while still retaining this property.

The probabilities of an upward or downward move did not enter into the valuation formula. Hence, we would obtain the same result if q depended on the current or past stock prices or on other random variables. In addition, u and d could have been deterministic functions of time. More significantly, the size of the percentage changes in the stock price over each period could have depended on the stock price at the beginning of each period or on previous stock prices.[18] However, if the size of the changes were to depend on any other random variable, not itself perfectly correlated with the stock price then our argument will no longer hold. If any arbitrage result is then still possible, it will require the use of additional assets in the hedging portfolio.

We could also incorporate certain types of imperfections into the binomial option pricing approach, such as differential borrowing and lending rates and margin requirements. These can be shown to produce upper and lower bounds on option prices, outside of which risk less profitable arbitrage would be possible.

Since all existing preference-free option pricing results can be derived as limiting forms of a discrete two-state process, we might suspect that two-state stock price movements, with the qualifications mentioned above, must be in some sense necessary, as well as sufficient, to derive option pricing formulas based solely on arbitrage considerations. To price an option by arbitrage methods, there must exist a portfolio of other assets which exactly replicates in every state of nature the payoff received by an optimally exercised option. Our basic proposition is the following. Suppose, as we have, that markets are perfect, that changes in the interest rate are never random, and that changes in the stock price are always random. In a discrete time model, a necessary and sufficient condition for options of all maturities and striking prices to be priced by arbitrage using only the stock and bonds in the portfolio is that in each period.

a. the stock price can change from its beginning-of-period value to only two ex-dividend values at the end of the period, and

b. the dividends and the size of each of the two possible changes are presently known functions depending at most on: (i) current and past stock prices, (ii) current and past values of random variables whose changes in each period are perfectly correlated with the change in the stock price, and (iii) calendar time.

The sufficiency of the condition can be established by a straightforward application of the methods we have presented. Its necessity is implied by the discussion at the end of section 3.19.20.21

[18] Of course different option pricing formulas would result from these more complex stochastic processes. See Cox and Ross (1976) and Geske (1979). Nonetheless, all option pricing formulas in these papers can be derived as limiting forms of a properly specified discrete two- state process.

[19]Note that option values need not depend on the present stock price alone. In some cases. formal dependence on the entire series of past values of the stock price and other variables can be summarized in a small number of state variables.

[20] In some circumstances, it will be possible to value options by arbitrage when this condition does not hold by using additional assets in the hedging portfolio. The value of the option will then in general depend on the values of these other assets, although in certain cases only parameters describing their movement will be required.

[21] Merton's (1976) model, with both continuous and jump components. is a good example of a

This round out the principal conclusion of this paper: the simple two -state process is really the essential ingredient of option pricing by arbitrage methods. This is surprising perhaps given the mathematical complexities of some of the current models in this field. But it is reassuring to find such simple economic arguments at the heart of this powerful theory.

Stock price process for which no exact option pricing formula is obtainable purely from arbitrage considerations. To obtain an exact formula it is necessary to impose .restrictions on the stochastic movements of other securities as Merton did or on investor preferences. For example Rubinstein 11976) has been able to derive the Black-Scholes option pricing formula under circumstances that do not admit arbitrage, by suitably restricting investor preferences. Additional problems arise when interest rates are stochastic, although Merton (1973) has shown that some arbitrage results may still be obtained.

References

Black. F. and M. Scholes. 1973, The pricing of options and corporate liabilities., Journal of Political Economy J. 637 654.

Brennan. MJ. and E.S. Schwartz. 1977, The valuation of American put options. Journal of Finance 32. 449 462..

Cox. J.e. and S.A. Ross. 1975. The pricing of options for jump processes. Rodney L. White Center Working Paper no. 2-75 (University of Pennsylvania, Philadelphia. PA).

Cox. J.e. and S.A. Ross. 1976. The valuation of options for alternative stochastic proces..c;es.

Journal of Financial Economics 3.145.166. Geske. R., 1979, The valuation of compound options, Journal of Finahcial Economics 7. 63-8l. Uarrison, J.M. and D.M. Kreps. 1979. Martingales and arbitrage in multi period securities markets. Journal of Economic Theory 20. 381-408.

Merton. R.C. 1973. The theory of rational option pricing. Bell Journal of Economics and Management Science 4. 141- 183.

Merton. R.C. 1976. Option pricing when underlying stock returns are discontimlous. Journal of Financial Economics 3, 125-144.

Merton. R.e.. 1977. On the pricing of contingent claims and the Modigliani-Millet theorem.

Journal of Financial Economics 5, 241 -250. Parkinson. M.. 1977. Option pricing: The American put. Journal of Business SO. 21-36. Rendleman. RJ. and BJ. Bartler. 1978. Two-state option pricing, Unpublished paper (Graduate School of Management. Northwestern University. Evanston. IL).

Rubinstein. M.. 1976. The valuation of uncertain income str.eams and the pricing of options. Bell journal of Economics 7. 407 u425.

Sharpe. W.F.. 197R. Investments (Prentice-Hall. Englewood Cliffs. NJ).

ARTICLE ON BANC ONE CORPORATION

Banc One Corporation:

Asset And Liability Management

On November 15, 1993, Dick Lodge, Banc One Corporation's (Banc One's) chief investment officer (CIO), gathered his notes and headed for a meeting with John B. McCoy, Banc One's chairman and CEO. On the way, he recalled the lunchtime conversation on the golf course six weeks earlier, during which McCoy had first voiced concern over Banc One's falling share price-from a high of $483/4 in April 1993 to just $36 3/4 (see Exhibit 1). McCoy attributed the decline to investor concern over Banc One's large and growing interest rate derivatives portfolio. During their discussion in September, McCoy had asked Lodge, who was responsible for managing the bank's investment and derivatives portfolio, to think about ways to deal with this problem.

McCoy had been prompted into action not only by the continued price decline, but also by the comments of equity analysts who covered Bane One:

The increased use of interest rate swaps is creating some sizable distortions in reported earnings, reported earning assets, margins, and the historical measure of return on assets. . . Were Bane One to include [swaps] in reported earning assets, the adjusted level would be 26% higher than is currently reported. . . Given its large position in swaps, Bane One overstates its margin by 131% [and its! return on assets in excess of 0.20%. . . Adjusted for [swaps], Bane One's tangible equity-to-asset ratio would decline by 1.55%.[3]

Banc One's investors are uncomfortable with so much derivatives exposure. Buyers of regional banks do not expect heavy derivatives involvement. . . Heavy swaps usage clouds Bane One's- financial image [and is] extremely confusing. . . It is virtually impossible for anyone on the outside to assess the risks being assumed. [4].

What made this situation more perplexing was that Banc One already had attempted to pre-empt concern over its growing derivatives portfolio. Along with its second-quarter results, it distributed a booklet detailing its asset and liability management policies and describing its derivatives portfolio, which had grown during the quarter from $23.4 billion to $31.5 billion in notional principal[5]. Lodge and others believed that the information in the booklet would help assuage any investor's concerns. Yet, given these kinds of comments from the analysts, the message was clearly not getting through.

In Lodge's mind, there was a simple explanation for the large size of Banc One's derivatives portfolio: swaps were attractive investments that lowered the bank's exposure to movements in interest rates. Why the market was penalizing Banc One for something that *reduced* its exposure to risk remained a mystery to him. Earlier in the year, Lodge had expressed his puzzlemen(to a reporter: "Why in the world more banks don't look at interest rate swaps. . . I don't know. It's not an esoteric phenomenon anymore. [6]

Nevertheless, he knew that *McCoy* attributed the decline to the derivatives portfolio and wanted to discuss alternatives for dealing with the situation.

Banc One Corporation[7]

Banc One Corporation, headquartered in Columbus, Ohio, truly epitomized the spirit of regional banking. With $76.5 billion in assets, it was the largest bank holding company based in Ohio and the eighth largest in the country. Unlike the more traditional bank holding company structure, in which the parent corporation controlled subsidiary banks, Banc One had a three-tiered organizational structure operating across 12 states. The parent, Banc One Corporation, controlled 5 state bank holding companies (in Arizona, Indiana, Ohio, Texas, and Wisconsin), which in turn owned. 42 subsidiary banks, or "affiliates." Through its Regional Affiliate Group, Banc One owned another 36 subsidiary banks-for a total of 78 banking affiliates. In addition to its banking affiliates, Banc One controlled 10 nonbanking organizations in various businesses ranging from insurance to venture capital to data processing.

For its banking business, Banc One had a very well defined, three-pronged strategy: concentrate on retail and middle-market commercial customers; use technology to enhance customer service and to assist in the management of banking affiliates; and grow rapidly by acquiring profitable banks.

Since 1969, it had completed 76 acquisitions involving 139 banks. In just the 10 years since 1982, it had completed 50 acquisitions, making it one of the top 10 corporate acquirers in the country.8 As of November 1993; Banc One had ten pending acquisitions that would bring an additional S9 billion in assets to the corporation. One of the largest pending acquisitions was Liberty National Bancorp, a bank holding company in Louisville, Kentucky with $4.7 billion in assets.

This deal highlighted many of the principles that guided Banc One's acquisitions. The target, Liberty National, had a strong retail focus, had a solid management team, and was the market leader. In addition, the deal was structured like most of its previous acquisitions: it would be accounted for as a: pooling of interests, be paid for with stock, and consist of a tiered offer that depended on the value of Banc One's stock price. The terms of the Liberty National Bancorp deal were as follows:

Banc One's Stock Price	Ratio of Banc one's Shares to Liberty National's Shares
Under $41.57	0.8421
$ 41.57 to $ 44.00	$ 35.00 worth of stock
above $ 44.00	0.7954

As of mid-November, Banc One's stock was trading near the "walkaway" price of $34.55. If it was below $34.55 in the second quarter of 1994, when the deal was expected to be consummated, one of two things would happen. Either Liberty National would cancel the deal or Banc One would end up using stock that it felt was undervalued to pay for the deal. Thus, a low stock price would either bring Banc One's acquisition program to a halt or cause it to violate one of its cardinal rules of acquisitions: acquisitions should not be dilutive. According to John McCoy, Banc One has "very strong pricing discipline. We just don't do dilutive acquisitions".[9] William Boardman, an Executive Vice President at Banc One, elaborated: "When we talk to prospects, we tell them we want the deal. to be non-dilutive when we do it, but that we also want it to be non-dilutive next year, and the year after that. Basically, what that means is that you have to grow your earnings at the same rate we're [Banc One] growing our earnings."[10]

While a strict set of principles guided Banc One's acquisition strategy, another well-defined set of principles guided its operating strategy. Internally, the operating strategy was known as the "uncommon partnership," which described the relationship among the affiliate banks and the various parts of the corporation. According to this partnership, the corporation decentralized the "people" side of the business and centralized the "paper" side. To capture the local knowledge of customers and markets, Banc One retained existing management in acquisitions and gave affiliate managers complete autonomy in running their banks. In contrast, Banc One centralized all of the affiliates' data processing, record keeping, and back office operations. This centralization fit well with Banc One's growth strategy. According to Boardman, "Growing just to become larger is not part of our strategy. Growing our economies of scale is."[11] The centralization of operations also capitalized on Banc One's vast experience with computer systems.

Over the years, Banc One had invested heavily in technology and information systems to support the uncommon partnership. Starting at the top with John B. McCoy, there was the belief that information was critical to running such a decentralized organization. One of the most important jobs of Banc One was to gather information from and disseminate it to the affiliates using the Management Information and Control System (MICS). This database tracked financial, productivity, and performance data for all affiliates. Every month, affiliates entered into the database their results and their revised budgets. In return, all affiliate presidents received a one-inch-thick report containing comparative statistics ranking all affiliates. The objective of this system was to encourage friendly competition among banking affiliates and to encourage managers to share information about effective banking products and practices.

Although it was an extremely complicated and highly decentralized organization, Banc One had one of the best financial track records of any bank in the country. Compared with the financial performance of the country's 25 largest bank holding companies in the decade since 1982, it had the highest average return on assets, the highest average return on equity, and the highest ratio of common equity to assets. Even more incredible was that Banc One had a string of 24 years of increasing earnings per share; none of the other large banks had a string of more than 7.[12]

Exhibit 2 summarizes Banc One's operating results and financial performance during the period 1983 to the third quarter of 1993.

Asset and Liability Management

A typical U.S. bank's liabilities consisted of floating-rate liabilities (such as federal funds borrowings) and long-term fixed-rate liabilities (such as certificates of deposit, or CDs). Assets included floating-rate assets (such as variable-rate mortgages and loans, as well as floating-rate investments) and long-term fixed-rate assets (such as fixed-rate mortgages and securities). Asset and liability management involved matching the economic characteristics of a bank's inflows and outflows. For example, a bank could match the maturity of its assets and liabilities. It also could look at the duration, the contractual fixed/floating nature of its commitments, or an estimate of the period in which its commitments would be repriced in response to changes in market rates as the basis upon which to judge just how well it was matched.

Banks' needs to match assets to liabilities arose from their strategic decisions regarding interest rate exposure. A bank could engineer its assets and liabilities to ensure that its earnings or market value would be unaffected by changes in interest .rates. Alternatively, a bank could adjust its portfolio of assets and liabilities to profit when rates rose, but lose when they fell. It could also position itself to gain when rates fell, and lose when they rose. The selection of interest rate exposure was a major policy decision for financial institutions.

In practice, banks typically had relative

In practice, banks typically had relatively more long-term fixed-rate liabilities (such as CDs) than they had long-term fixed-rate assets (such as loans). To make up for this shortfall, banks that wished to match assets and liabilities complemented their loan portfolios with fixed-rate investments commonly called balancing assets, such as Treasury securities. By adjusting the characteristics of the balancing assets, a bank could better match its assets to its existing liabilities.

As chief investment officer of Banc One, Dick Lodge managed the firm's portfolio of balancing assets. His staff of approximately 100 people, with 12 engaged in asset and liability management activities, measured the degree to which the bank's assets and liabilities were matched and made profitable investments consistent with the bank's policy of managing its interest rate exposure. Specifically, they had an official mandate to (1) invest funds in conventional investments and derivatives to conserve the funds' principal value yet provide a reasonable rate of return; (2) keep enough funds in liquid investments to allow the bank to react quickly to demands for cash; (3) control the exposure of Banc One's reported earnings to swings in interest rates; and (4) achieve these objectives without unnecessarily increasing the bank's capital requirements.[13]

In carrying out this mandate, Banc One used investments and derivatives as substitutes for one another. For example, if it wanted to increase its share of fixed-return investments, it could sell a floating-rate investment (or borrow at a floating rate) and use the proceeds to buy a three-year fixed-rate Treasury note. The initial net outflow of these two transactions would be zero, but the transactions would increase the relative magnitude of the bank's fixed-rate portfolio. Alternatively, Banc One could enter into an interest rate swap in which it paid a floating rate of interest and received a fixed rate in return. The initial net outflow of such a swap also would be zero. As in the first example, such a transaction would increase the bank's fixed-rate inflows and reduce its periodic net floating-rate inflows. Because the security transactions and the swap produced similar interest rate exposure, they had to be compared on other dimensions, such as yield, credit risk, capital requirements, transaction costs, and liquidity.

Defining and Measuring Interest Rate Exposure

Banc One, like other banks, defined its exposure to interest rate risk by calculating its earnings sensitivity, or the impact of interest rate changes on reported earnings. For example, if a gradual 1% upward shift in interest rates during the year increased that year's base earnings by 5%, the bank would have an earnings sensitivity of 5%. If earnings sensitivity was positive, the bank was said to be asset sensitive (i.e., the interest rate on assets reset more quickly than liabilities, resulting in increased income if rates rose). If earnings sensitivity was negative, the bank was said to be *liability sensitive* (i.e., liabilities reset more quickly than assets, resulting in a decrease in income if rates rose). If the bank had a 0% earnings sensitivity, then an upward or downward shift in interest rates would have no effect on its earnings.

Like many banks, Banc One's basic portfolio (excluding its balancing assets) was asset sensitive. Its asset sensitivity arose because a large proportion of its assets, such as commercial loans, were indexed to the prime rate and therefore varied contractually with market rates. However, the bank's liabilities included mostly fixed-rate items such as fixed-rate CDs as well as "sticky-fixed" savings and demand deposits whose rates changed much more slowly than market indices. Banc One's relative overabundance of fixed-rate liabilities would make its earnings increase as rates rose. This natural asset-sensitivity was exacerbated by its acquisition program because many of the banks it acquired were highly asset sensitive.

Over the years, Banc One's evolving program to measure interest rate risk mirrored best practice in the U.S. banking industry. Prior to the 1980s, the bank did not precisely measure its exposure to changes in interest rates. Instead, it generally avoided investing in longer-maturity securities, feeling that these investments could add undue risk to the liquidity of its investment portfolio. By the early 1980s, it had become clear to Banc One's management that measuring interest rate risk was a critical task. The second oil shock of the 1970s had increased the level and volatility of interest rates. For example, the prime rate soared to more than 20% in late 1980, twice the average for the 1970s and four times as large as the average in the 1960s. In 1980 alone, the prime rose to 19.8% in April, fell to 11.1% in August and rebounded to more than 20% at the close of the year. To determine the bank's exposure to interest rate movements in this new, more volatile interest rate environment, Banc One began measuring its *maturity gap* in 1981.

Maturity gap analysis compared the difference in maturity between assets and liabilities, adjusted for their repricing interval. Repricing interval referred to the amount of time over which the interest rate on an individual contract remained fixed. For example, a three-year loan with a rate reset after year one would have a repricing-adjusted maturity of one year. Banc One grouped its assets and liabilities into categories, or "buckets," on the basis of their repricing-adjusted maturities (less than 3 months, 3 to 6 months, 6 to 12 months, and more than 12 months). The maturity gap for each category was the dollar value of assets less liabilities. If the bank made short term floating-rate loans funded by long-term fixed rate deposits, it would have a large positive maturity gap in the shorter categories and a large negative maturity gap in the longer periods.

The maturity gaps could then be used to predict how the bank's net interest margin (the difference between the weighted average interest rate received on assets and the weighted average interest rate paid on liabilities)-and therefore earnings-would be affected by changes in interest rates. For example, if interest rates dropped sharply, a large positive maturity gap for the short maturity buckets would predict a drop in interest income and. therefore earnings, because the bank would immediately receive lower rates on its loans while still paying higher fixed rates on its deposits.

Unfortunately, implementing the initial maturity gap measurement program was extremely time consuming. By the time each gap report was collected from the affiliates, consolidated, and analyzed, the information was dated. Lodge himself constructed the first gap management report in 1981, and it took almost a year to complete.

In 1984, Banc One began using asset and liability simulations as a more accurate method measure its exposure to interest rates. By using exact asset and liability portfolios rather than grouping each asset or liability according to its repricing interval, Banc One was able to measure how interest rate changes would affect earnings. To do so, it created an "on-line balance sheet" that contained upto-date information on its assets and liabilities, which complemented the MICS process. The key features of each contract, including principal amounts, interest rates, maturity dates, and any amortization schedules of assets and liabilities, were recorded. Then, Banc One used historical data to estimate such items as the maturity of demand-deposit (checking) accounts, the speed with which its bank managers would reprice deposits and loans in response to interest rate shifts, and the rate at which its borrowers might refinance fixed-rate loans if rates dropped..

Once the model was complete, Banc One could simulate how any shift in interest rates would affect its balance sheet and earnings, as well as run sensitivity analyses on its assumptions. Although the model had been refined since 1984, it served as the basis for measuring the bank's interest rate risk and senior management reviewed its predictions monthly. In 1993, this on-line balance sheet was redesigned to include a monthly down-load of each of over 3 million loans or deposits, that is, a discrete asset and liability database on each customer that included prepayment, optionality, and convexity estimates.[14]

Investments for Managing Interest Rate Exposure

Banc One's evolving sophistication in managing interest rate exposure mirrored its sophistication in measuring it. In the early 1980s, it managed its exposure to interest rate risk by adding balancing assets to its investment portfolio until it felt it had enough fixed-rate investments to offset its fixed-rate liabilities. In 1981, 13% and 21% pf Banc One's earning assets were money market investments and longer-term securities, respectively. Initially, Banc One invested in short- and medium-term U.S. Treasuries and high-quality municipal bonds. Municipal bonds were an especially attractive investment because prior to 1986, banks could deduct 800% of the interest expense incurred on monies raised to buy them. Because the income earned on the bonds was free of state and federal taxes, banks could enjoy a large after-tax spread on their leveraged municipal bond investments.

In 1983, Banc One began using interest rate swaps as part of its investment portfolio. Originally, swaps were used to lock in high after-tax yields on municipal securities. By buying the municipal bonds, Banc One .received an after-tax yield of 9.50%. By then entering into an interest rate swap in which it paid a fixed rate of 7.00% and received the London Interbank Offered Rate (LIBOR), a commonly used floating-rate index, it ended up with a net position of receiving LIBOR + *2.50010.* The bank's net cash flow *from* the investment and swap resembled a floating rate investment with an above-market yield. During the course of 1983 and 1984, Banc One became increasingly comfortable with the use of swaps as a tool to tailor individual investments to suit its needs.

In 1986, Congress passed the Tax Reform Act, which eliminated *for* banks the deduction of interest expense on the financing *for* municipal bond investments.[15] Banks turned to other investments that would provide the same high yield they had grown accustomed to receiving. Banc One replaced many of its municipal investments with mortgage-backed securities (MBSs), which were fixed-income investments whose payment stream was backed by pools of mortgage loans and which were typically guaranteed by the federal government. MBSs provided a slightly lower promised after-tax yield than did municipal bonds and carried an additional risk of prepayment. If interest rates *fell*, borrowers typically refinanced their mortgages by prepaying their existing mortgages. The owner of a pool of mortgages was forced to reinvest precisely when market yields were relatively low and was left with a submarket yield when rates *rose.*

In 1983, Wall Street created anew type of mortgage security: the CMO, or collateralized mortgage obligation. CMOs took a pool of mortgage loans and carved the principal and interest outflows into a set of different securities, or trenches. The trenches differed from one another only in their priority for repayments of principal. For example, the first tranche of a CMO would receive all of the mortgage prepayments until its principal was returned to its holders. At that point, the second tranche would begin to receive prepayments until its principal was fully paid out, and so on. With a large pool of mortgages, investors could statistically estimate the likely speed of prepayment and therefore the likely time at which each tranche would be fully paid down and stop paying interest. Each tranche paid a different yield to compensate *for* the various amount of prepayment risk a buyer faced, as well as *for* the different average life of the investments. By investing in CMOs, Banc One could still receive the high yields associated with mortgage securities, assuming it was comfortable with the prepayment risk it would bear. In 1993, Banc One had 54.5 billion invested in CMOs, or about a third of their investment portfolio. Earlier in the 1980s, as much as two-thirds of their investment portfolio was held in CMOs.

Swaps as Synthetic Investments

After using swaps in the mid-1980s to tailor cash flows of individual municipal investments, Bane One realized that it could also use swaps as a proxy *for* some of its conventional fixed-rate investments. Instead of investing in medium-term *U.S.* Treasury obligations, it could simply enter into a medium-term receive-fixed swap and put its money into short-term floating-rate cash equivalents. There were several advantages of this "synthetic investment" over conventional investments.

First, the swap greatly improved the bank's liquidity. Banks need cash to accommodate customer withdrawals and to repay existing liabilities, such as CDs, as they mature. Investing in long-dated securities could increase a bank's yield, but if the bank needed to raise cash suddenly, these investments might not be easily liquidated or their liquidation might expose the bank to a large loss in principal. With a swap, the bank could invest in short-term, highly liquid securities with stable principal values. By layering a receive-fixed swap onto this investment, the bank could obtain the economics of the longer-term investment, while still enjoying the high liquidity of the short-term instrument.

Second, unlike investments and borrowings, swaps were off-balance-sheet transactions. If Bane One were to buy a fixed-rate bond and sell a floating rate security, both would appear on its balance sheet, and the spread between the two would appear as income. However, if it were to enter into a receive fixed swap with the same cash flow implications, the swap would not appear as either an asset or liability, but would be disclosed only in footnotes to the financial statements. Yet the current net income or loss from the swap transaction still would appear on its income statement. This accounting treatment would tend to overstate traditional profitability measures such as a bank's return on assets in comparison to the identical securities transactions.

Finally, in comparison to a conventional securities investment, swaps could also reduce the amount of capital needed to meet regulatory requirements. These minimum capital requirements grew out of an international agreement, the Basle Accord, signed by the central bankers of the major industrialized countries. In agreement with the Accord, U.S. banking regulators implemented risk-based capital standards beginning in December 1990. The new regulations dictated the amount of capital banks needed to hold as a function of their total risk-based assets.[16] As of year-end 1992, u.s. regulators raised the minimum capital levels and strengthened their power to close institutions that failed to meet these minimums.

Stricter capital standards led banks to prefer assets with lower capital requirements, all else being equal. Some observers attributed the rising growth in bank investments in Treasury securities to their zero risk weighting in the calculation of risk-adjusted assets. Under the capital guidelines, swaps contributed little to the risk-adjusted assets against which the bank had to hold capital.[17] Were a bank to create exposure similar to the swap using securities (other than U.S. Treasury securities), its need to hold capital would be *20010* to 100% of the principal value of the assets.[18]

During the late 1980s, Banc One began replacing many of its maturing conventional investments with synthetic investments. As part of this trend, it began to investigate whether it could create a synthetic CMO, which would have the advantages of other swaps, yet deliver the risk/return characteristics of CMO investments. Specifically, a synthetic CMO would allow Banc One to enjoy high yields in exchange for taking on prepayment risk. After a few false starts and discussions with various investment banks, Banc One and its counterparties developed a product called Amortizing Interest Rate Swaps (AIRS).[19]

Because AIRS replicated investments in mortgage securities, they needed to have similar prepayment features. With low interest rates, consumers prepay their mortgages, and mortgage investors receive back their principal. In the AIRS, the notional amount of the swap would be reduced or amortized if interest rates fell. As interest rates declined, the AIRS would amortize faster, thereby leaving the bank to reinvest just when market yields were low. Likewise, when interest rates increased, the maturity of an AIRS would end up longer than expected, thereby leaving the bank with a below-market yield on its investment. In early AIRS, the amortization of the notional principal balance was tied to the performance of a particular pool of actual mortgages, but with later AIRS, the amortization schedule was set by a formula. Exhibit 3, panel A, gives the terms for the latter type of AIRS.

As synthetic investments, AIRS produced attractive yields. In these transactions, Banc One would receive a fixed rate of interest and pay LIBOR. In 1993, this fixed rate, called a swap spread, was perhaps 120 basis points over a Treasury security of the same maturity. In comparison, the bank could buy a comparable CMO and receive a yield of 100 basis points over Treasuries. If Banc One was to enter into a standard (non-amortizing) swap of the same term, it might receive a fixed rate of 20 basis points over Treasuries.

With Banc One's mortgage portfolio as well as its investments in CMOs and AIRS, prepayment risk complicated the task of measuring interest rate risk. The embedded options that Banc One sold to its mortgage borrowers, certain depositors, and to its swap counterparties made its earnings sensitivity nonlinear. With a rise in rates, the earnings from its fixed-rate investments would not change. However, a drop in rates which precipitated prepayments of mortgages or amortization of the AIRS forced the bank to reinvest the early repayment of principal at the lower market rates.. Furthermore, steep rate drops typically increased the rates of prepayment or amortization. For example, though earnings might drop 1% for a 1% increase in rates, a 2% increase in rates might reduce earnings by 3% or 4%, not 2%.

Exhibit 3

Panel A : Amortizing Interest Rate Swap (Airs) September 1993

REPRESENTATIVE SWAP

TRANSACTIONS

Notional amount ………….. $ 500 million.

Final maturity……………… 3 years (if not amortized early)

Payment Frequency……….. Quarterly

Banc one pays………………3 – month LIBOR (3.25% at initiation of swap)

Banc one receives ………….. 4.5%

Lock out period …………….1 year (During the lockout period, there is no amortization of swap)

Cleanup provision ……….. 10% of original national amount (If the notional amount falls to $50 million or less through amortization, the swap is cancelled)

Amortization schedule…….. Each quarter, after the lockout period, the notional principal of the swap is reduced by the following amount for the following quarter, depending on the level of interest rates.

If 3 – month LIBOR	Notional Principal Amount	Average Life of swap
Stays at 3.35% or falls	Completely amortized	1.25 years
Rises to 4.35%	Reduced by 31%	1.75 years
Rises to 5.35%	Reduced by 10.5%	2.50 years
Rises to 6.35 or higher	Not reduced	3.25 years

Panel B: Libor – Prime Basis Swap

Notional amount ……… $ 200 million

Final maturity ………… 4 years

Payment frequency …… Quarterly

Banc One pays ………… Daily average prime rate – 270 basis points (At initiation, prime was 6%)

Banc One receives …… 3-month LIBOR (subject to caps) (At initiation, 3 month LIBOR was 3.375%)

Caps ………………… In no quarterly period can the rate Banc One receives exceed 25 basis point over the rate received in the prior quarter.

Swaps as a Tool for Risk Management

Banc One had a long-standing stated policy of "minimizing the impact of fluctuating interest rates on earnings and market values,"2O and in 1986, its senior management adopted guidelines for allowable earnings sensitivity. This first policy stated that earnings could not change more than 5% for a 1% immediate change in interest rates. Because Banc One was more asset sensitive than its policy would permit, the bank considered alternatives for adjusting its earnings sensitivity, finally using swaps as its solution.

Although in the past the bank had entered into pay-fixed swaps to transform the cash flows on its municipal investments, the exact opposite swap was required to shift it away from an asset-sensitive position and toward more liability sensitivity. By entering into an interest rate swap in which it paid a floating rate and received a fixed rate in return, it was as if the bank was incurring a floating-rate liability while investing in a fixed-rate asset. This combination would move the bank toward a liability-sensitive (or negative earnings-sensitive) position. If interest rates rose, the floating-rate payments on the swaps would increase the bank's interest expense while interest income remained constant, thus reducing earnings and producing liability sensitivity. As Banc One gradually enlarged its interest rate swap portfolio in the mid-1980s, its earnings sensitivity moved to within the specified 5% boundary. See Exhibit 4 for historical information on Banc One's investment portfolio, swap portfolio, maturity gap, and earnings sensitivity during the period 1988 through the third quarter of 1993.

EXHIBIT 4 BANC ONES INVESTMENT PORTFOLIO AND INTEREST RATE SENSITIVITY / 1998 THROUGH 1995 Q3($ MILLIONS)

	Investments							Swaps				
	Amount Outstanding				Gross Income Received			Gross	Maturity			
	Earnings		Short-term			Short term		Amount	Income	Gap	Earnings	
	Assets	Loans	Investments	Securities	Loans	Investment	Securities	Outstanding		Received.[b]	Measures[c]	Sensitivity
1988	$22,531	$17,325	$581	$4,625	$1,876	$28	$368	N/A	N/A	-6.67%	-1.00%	
1989	23,568	17,909	525	5,133		2,167	39	446	$3,299	$291	-3.59%	-1.00%
1990	26,680	20,363	628	5,272		2,303	58	441	3,231	292	-10.07%	-1.55%
1991	41,482	31,168	2,324	7,989		2,747	61	484	11,214	887	-7.33%	-2.30%
1992	54,766	39,142	1,740	13,884	3,872	86	870	10,492	766	-15.70%	-2.61%	
1993:Q1	61,807	45,361	1,382	15,064		1,159	11	231	14,132	240	-2.34%	-2.50%
1993:Q2	66,796	48,845	1,978	15,973		1,173	9	235	17,280	275	-2.65%	-2.60%
1993:Q3	68,116	50,105	1,217	16,794		1,189	10	216	22,515	335	-3.64%	-3.30%

a. Includes only receive-fixed swaps.

b. Notional volume of outstanding receive-fixed swaps multiplied by average fixed rate received on such swaps.

c. Maturity gap over the first one-year horizon as a percentage of earning assets, where maturity gap is defined as total assets with adjusted maturity of one year or less minus total liabilities with adjusted maturity of one year or less.

Sources: Banc One Corporation. Annual Reports. 10-Ks.

Because the swaps were designed to adjust the bank's *earnings* sensitivity, the greater its earnings, the more swaps it would need. Also, the more its natural earnings sensitivity strayed from the policy guidelines, the more swaps it would need. Both of these factors contributed to the subsequent growth in its swap portfolio. For example, in 1989, Banc One' acquired banks with $12 billion in assets from M Corp, a failed Texas bank. These banks were 23.4% asset sensitive when they were bought, far outside Banc One's policy target range and well above its then-slight liability sensitivity. To bring the new banks in line, Lodge had to enter into a large notional volume of swaps. The bank's continued acquisition strategy, as well as its earnings growth, would increase its need for swaps.[21]

Managing Basis Risk

Though synthetic investments reduced Banc. One's earnings sensitivity to overall shifts in interest rates, they created a heightened sensitivity to mismatches *between* floating-rate interest rates, or basis risk. Most of Banc One's floating-rate assets were based on the prime rate. However, most conventional interest rate swaps as well as its AIRS used three-month LIBOR as an index for floating-rate payments. LIBOR was an actively traded global market rate that changed daily. In contrast, the prime rate was an administered U.S. or local rate that changed infrequently at bankers' discretion. Because of these differences, the spread between the two rates changed dramatically over time. (See Exhibit 5 for a graph of prime and three-month LIBOR.)

For example, assume the baI1k entered into a swap in which it received 7% and paid LIBOR. Ignoring the difference between prime and LIBOR, it would effectively transform its prime-based floating-rate assets into fixed-rate investments paying 7%. However, if three-month LIBOR increased 150 basis points but prime was unchanged, Banc One would have transformed its prime-based floating-rate asset into a fixed-rate asset paying not 7% but 5.5%, and it would have created basis risk through its exposure to swings in the prime-LIBOR spread.

To counter this basis risk, Banc One entered into basis swaps that reduced the floating-rate mismatch (see Exhibit 3, panel B, for typical basis swap terms), In a basis swap, Banc One would pay a floating rate based on the prime rate and receive a floating rate based on three-month LIBOR. This contract would offset the spread differential between prime and three-month LIBOR. Using a basis swap in conjunction with an AIRS in which it paid LIBOR, Banc One could confidently transform prime-based floating-rate assets to fixed-rate investments.

Managing Counterparty Risk

The credit risk of investing in swaps differed from that of traditional investments, If Banc One bought a U.S. Treasury bond, for example, it would face no credit risk. However, if it entered into a swap transaction in which it received the fixed rate, it would be exposed to the default of its counterparty.

This credit risk was mitigated in three ways. First, the positive swap spread (i.e. "yields on swaps were higher than on Treasury securities") gave the

Exhibit 6 Banc One's Exposure To Its Major Swap Counterparties October 31, 1993 ($ In Millions)

	National Amount	Average Maturity	Mark to Market Exposure[a]	Collateral Posted[b]	Net MTM Exposure[c]	Potential Exposure[d]	Net Credit Exposure[e]
Bankers Trust	$12,142	1.77	$123	$132	($9)	$ 68	$ 59
Union Bank of Switzerland	6,976	1.87	49	49	0	92	92
Goldman Sachs	6,163	1.57	58	122	(64)	26	(38)
Lehman Brothers	4,058	2.32	16	81	(65)	26	(39)
Merrill Lynch	3,347	2.17	59	104	(45)	10	(35)

a. Mark to market exposure measured as the market value of swap positions with counterparty, A positive exposure indicates that Bane One's swaps have a market value greater than zero.

b. Collateral is posted in the form of cash or bank-eligible securities. A positive number indicates that Bane One's counterparties have deposited collateral with Bane One.

c. Represent. mark to marker (MTM) exposure less collateral posted by Bane One's counterparties.

d. The bank estimated its potential exposure if it experienced a large movement in interest rates relative to historical experience. Specifically using historic data, it calculated the distribution of interest rate moves Over 30 days. It then calculated how much it could lose. If rates moved in Bane One's favor, and if the size of the rare move was equal to a three-standard deviation change in rates. 99% of all rate moves would he within three standard deviations. So this measurement of its potential gains was considered a conservative estimate of the bank's credit exposure.

e. Represent. Bane One's potential loss less the collateral it currently has on hand. Source: Bane One Corporation.

bank a higher return to compensate for its credit risk. Second, in an investment, the bank's entire principal was at risk (if the issuer was not the u.s. government), whereas in a swap, only the net payment (fixed less floating) was at risk of default. Third, Banc One established strict policies for managing its counterparty exposure.

In all instances, its counterparties were rated no lower than single-A. To understand its potential exposure, Banc One continually monitored its mark to-market exposure to each counterparty. Its total exposure to any entity, whether through derivatives or direct lending, was limited by clear policy guidelines. In addition, to protect itself against the default of a swap counterparty, Banc One required its counterparties to post collateral, in the form of bank eligible securities or cash, against the bank's exposure.[22] Investment bank counterparties posted collateral at the initiation of the swap equal to Banc One's possible losses from an extreme one-month move in interest rates.[23] All counterparties were required to post additional collateral as the market value of the swap changed over time.[24] This practice meant that Banc One was not exposed to swap payments for which it did not have collateral, and were the counterpart)' to default, the mark-to-market collateral would allow the bank to enter into a new swap that was economically identical to the one that had defaulted. Banc One's counterparties-and its exposures to each-are shown in Exhibit 6. Banc One's collateral requirements were unique, as most large money-center banks and commercial banks were extremely reluctant to post any kind of collateral for swaps, regardless of the counterparty. Yet, because of the magnitude of its derivatives portfolio and because of its solid credit rating, Banc One was almost always able to secure such collateral agreements, even from AAA-rated counterparties.

Controlling the Asset and Liability Management Process

Banc One's careful handling of counterparty risk was indicative of its long-standing, well-defined investment policies. In late 1993, the investment policies of many banks (including Banc One), and especially their use of derivatives portfolios, came under public scrutiny.

In mid-1993, a consortium of leading financial service firms, known as the Group of Thirty, released a report in which it recommended a set of practices that all derivatives dealers and users follow to ensure that these instruments were used prudently. This report was commonly seen as a proactive effort at self-regulation to fend off governmental regulation of derivatives. Later that year, in October, the U.S. Comptroller of the Currency, the regulator of national banks, issued its own set of guidelines for the use of swaps. The guidelines focused on the role of senior management and boards of directors in ensuring that users of swaps acted safely. The report charged banks with managing market risk, counterparty credit risk, liquidity risk, and operations and systems risk while remaining mindful of the impact of swaps on the banks' capital base and accounting. Politicians seized on the issue and made their own statements concerning the swap market. The statements of the industry, regulators, and politicians pushed the banking sector's use of derivatives onto the front pages of leading newspapers and made the issue, one of general interest.

This newfound .interest in the management of derivatives positions came as no surprise to Banc One. For years, senior management had made the prudent use of derivatives and ocher investments, as well as management of its assets and liabilities, a top priority. Its Asset and Liability Management Committees (ALCOs) were responsible for establishing and implementing policies relating to asset and liability management. The process was governed by a 70page policy document, updated in April 1993, which outlined an exact system of control and oversight of the bank's asset and liability management policies, including its management of swaps, an integral part of its investment portfolio. The ALCO process was a system for consistently managing interest rate risk, credit risk, funding risk, and capital adequacy. A committee of the most senior bank executives reviewed and ratified major investment decisions, recommended changes to existing policy, and monitored compliance with policy guidelines.

The ALCO process consisted of regular meetings at several levels of the bank. Affiliate banks reviewed their cash position and funds management activities daily. For each state, asset and liability committees were established to monitor that state's activities. At the corporate level, three committees met weekly or monthly to monitor and oversee the overall asset and liability system: the corporate funds management activity committee; the working ALCO committee, which included Lodge, McCoy, and many other senior executives; and the corporate ALCO committee, which included the working ALCO as well as the chairmen of Banc One's holding companies and its chief credit officer. The operation of the MICS system made timely and appropriate information available to each committee.

All policy decisions regarding Banc One's earnings sensitivity were made at the corporate level. Furthermore, the firm's investment activities, including both securities and swaps, were executed at the corporate level by CIO Dick Lodge and his group. Thus, the affiliate and state ALCO groups monitored local deposit and lending activities and their impact on the units' liquidity and interest rate exposure. Corporate ALCO activities overlaid investments and derivatives onto the aggregated activities of the local banks in order to manage the bank's overall exposure.

When it was established in 1986, the bank's policy was to stay within a 5% earnings sensitivity boundary for an *immediate* 1% shock to interest rates. However, Lodge had recently persuaded the working ALCO committee that such a shock was unrealistic. He believed the committee should instead focus on the impact of *agradual* 1% in the level of interest rates during the year (i.e. rates would slowly rise 1%, so that on average they would have risen *112%).* The working ALCO committee agreed to this change, and it also set a new boundary for the bank at 4% sensitivity. In addition, the committee set other guidelines:

Earnings Sensitivity	Policy	Nov. 1993 Banc One Position
1st-year impact for a +1% rate change	(4.00)%	(3.30)%
1st-year impact for a +2% rate change	(9.00)%	(8.00)%
1st-year impact for a +3% rate change	(1;.00)%	(13.20)%
2nd-year impact for a + 1 % rate change	(4.00)%	(1.30)%
2nd-year impact for a + 2% rate change	(9.00)%	(7.90)%
1st-year impact for a -1% rate change	(4.00)%	4.00%

Within these strategic guidelines, Lodge was permitted, with the working ALCO group's approval, to make tactical decisions on exactly what the bank's earnings sensitivity should be. Although there were several guidelines and Lodge had to comply with each one, both he and the ALCO groups focused mainly on the first-year impact of a gradual 1% change in rates because they believed it was unlikely that interest rates would change much more than 1% in the coming year.

a. Average yield received on investment portfolio (excluding swaps J. For projected period assumes no new investments made.

b. Average scale received on receive-fixed swap portfolio. For projected period assumes no new positions added.

Source: Bane One Corporation.

In November 1993, if it did not have its $12 billion in fixed-rate investments and $22 billion in receive-fixed swaps, the bank' would have been 13% asset sensitive. With them, it was positioned to be 3.3% liability sensitive. This conscious decision to be modestly liability sensitive was the bank's strategic exposure to interest rates. *As* Lodge explained, "Banks are paid to be liability sensitive," meaning that tDe yield curve was almost always upward-sloping. By having a controlled amount of long-term, fixed-rate, income-producing assets exceeding its short-term, floating-rate liabilities, the bank could earn the interest differential as long as the yield curve remained upward-sloping and did not shift up dramatically. However, this net position left the bank liability sensitive as a rise in rates would reduce its income.

Although a sudden rise in rates would depress the bank's earnings, the investment portfolio was set up so that this exposure was controlled. Specifically, the swaps in place were level over the next year, but would virtually all mature within two years. Thus, if the bank did not add new swaps to its position, its existing swaps would fall to $17.5 billion by year-end 1994 and $3.6 billion by year end 1995. Its projected earnings sensitivity would drop to -.2% by the end' of 1994, effectively making its earnings unaffected by interest rate swings, and the bank would be asset sensitive by 1995. See Exhibit 7.

Although the bank focused primarily on the impact of interest rates on its earnings, the ALCa committee also examined the effect of interest rates on the value of the firm and its common equity. The asset and liability database allowed it to measure the duration of assets and liabilities. Lodge's figures for the bank's key duration measures,[25] as of September 30, 1993, were 1.73 years for on- and off-balance sheet assets and 1.51 years for its liabilities. Because the difference between assets and liabilities was a residual equity account, Lodge could also calculate a rough duration of equity (by weighting each category by its total dollar amount). *As* of September 1993, residual equity had a duration of +4.00 years. For each 1% rise in rates, this duration measure suggested that Banc One's equity value would drop by 4.0%. As interest rates rose, its slightly longer duration asset base would decline in value faster than its shorter duration liabilities, leading to a magnified drop in the market value of its equity.

As of September 1993, Banc One had $37.7 billion in notional volume of interest rate, swaps on its books. Both Lodge and McCoy felt that the bank had drawn some of its unwanted attention because its swap portfolio had grown so dramatically. One analyst identified Banc One as having the second largest growth in an existing swap portfolio of all regional banks. At the end of 1990, Banc One had only $4.7 billion in swaps on its books. This figure had grown to 513.5 billion at the end of 1991 and $21.0 billion at the end of 1992. Looking forward, Banc One saw continued growth in its swap portfolio as long as its earnings grew, it continued to acquire banks that were more asset sensitive than itself, and it faced an upward-sloping yield curve.

Disclosure

As of November 1993, the Financial Accounting Standards Board (FASB) required minimal disclosure of the details of a company's swap portfolio because swaps were classified as off-balance-sheet items. Generally, the total notional volume of swaps was reported as a footnote to reported financial statements. Under accounting guidelines, though, notional volume had EO include *all* swaps , regardless of their purpose or whether they offset one another. Thus, if Banc One entered into a $100 million receive-fixed swap and then a $100 million basis swap to adjust the floating-rate index it paid, the swaps would be reported as $200 million of notional amount, even though they economically replicated only $100 million of a fixed-rate investment. Likewise, if it entered into a $100 million pay-fixed swap and then entered into an exactly offsetting receive fixed swap, it would report $200 million in swaps.

Even though FASB required minimal swap disclosure, Banc One had voluntarily disclosed additional information, consistent with its reporting policies. In addition to reporting the total notional volume of swaps on its books, it reported the unrealized net gain or loss on its swap portfolio. Banc One's disclosures of its swaps activities for 1993 are shown in Appendix 1.

The Meeting

As Banc One's earnings grew, so too did its swap position. With its growing swap portfolio, it caught the attention of bank analysts. Some applauded the bank's use of swaps to manage its interest rate exposure. Other-more vocal-analysts were critical, accusing Banc One of using swaps to inflate earnings, overstate capital ratios, and offset declines elsewhere in the bank. These critics saw the rapidly growing swap positions as heading out of control. One analyst was quoted as saying of the bank's swap activities, "Does that look like hedge activity? They use this stuff to keep the game going." A few analysts had downgraded the stock.

Though it was impossible to pin the recent decline in Banc One's stock price solely on its growing derivatives portfolio, both insiders and outsiders felt that the $10 drop in its stock price was due in large part to the market's reaction to the bank's use of derivatives. One analyst supportive of the company wrote:

> One likely reason for the price weakness is a recent focuses on Bane One's liberal use of derivatives to achieve its asset/liability management goals. Since derivatives are relatively new financial instruments, and since their use requires a high degree of financial sophistication and quantitative expertise, there is an understandable aversion to them on the part of many investors.. .Although (Bane One's swap position) is a large notional amount for a regional bank, we think Bane One's use of derivatives has been prudent.[26]

As the meeting between McCoy and Lodge began, McCoy voiced his concern about Banc One's falling stock price. From his perspective, he and Lodge faced a dilemma. On the one hand, he felt that swaps were hurting shareholder value because the investment community did not understand how they were being used. On the other hand, he believed that they were an invaluable tool in managing risk. Given the distance between his beliefs and. what he was hearing from the market, he wondered what, if anything, the bank should do.

In an attempt to answer this question, McCoy and Lodge discussed three possible options. First, they could do nothing and hope that Banc One's stock price would recover over time as investors realized that derivatives were actually helping the bank manage interest rate and basis risk. Second, they could abandon or severely limit their derivatives portfolio. Third, they could attempt to educate investors about how they used derivatives. Their most recent quarterly disclosure gave the market a great deal of data on the bank's swap portfolio, but perhaps even more information might dispel the misconceptions. What information would the market want to see? And how could Banc One credibly present it so as to convince its skeptics and educate swap novices? Perhaps analysts would understand Banc One's ALCO process and use of swaps if they could compare the bank to a hypothetical Banc One that had no swaps or no investments. In preparation for the meeting, Lodge and his staff prepared a set of analyses showing this comparison (see Appendix 2).

None of the alternatives was riskless. Doing nothing might give the impression that the bank was hiding something, thereby confirming investors' worst suspicions. If it caused Banc One's stock price to stay low or fall even further, the bank's ability to continue its stock acquisitions would be jeopardized. Eliminating its derivatives portfolio would leave the bank with ,greater interest-rate exposure and few tools to manage it. Disclosing even more information was not a guaranteed solution. In drawing even greater attention to its derivatives portfolio, the bank might raise investors' concerns or increase their confusion.

Appendix 1: Banc Ones 1993 Disclosure of Its Interest Rate Management Activities (10-q-fillings)

Panel A: 1993 First Quarter

BANC ONE manages interest rate sensitivity within a very small tolerance through the use of off-balance sheet interest rate swaps and other instruments, thereby minimizing the effect of interest rate fluctuations on earnings and market: values. The use of swaps resulted in BANC ONE being slightly liability-sensitive at March 31, 1993, countering the natural

Panel B: 1993 Second Quarter

BANC ONE manages interest rate sensitivity within a very small tolerance through the use of off-balance sheet interest rate swaps and other instruments, thereby minimizing the effect of interest rate fluctuations on earnings and market values. The use of swaps resulted in BANC ONE being slightly liability-sensitive at June 30, 1993, adjusting the natural tendency to be asset-sensitive. Swaps increased interest income by $59 million and $113 million for the three and six month periods ending June 30, 1993 as compared to $46 million and 595 million for the same periods in 1993. Swaps decreased deposit and other borrowing cost by $48 million and 596 million for the three and six month periods ended June 30, tendency to be asset-sensitive. The use of swaps to manage interest rate sensitivity increased interest income by $54 million and 550 million, and decreased interest expense by $47 and $34 million in the first quarter of 1993 and 1992, respectively. The notional amount of swaps increased from $8.3 billion to $23.4 billion from March 31, 1992 to March 31, 1993.

1993, compared to decreases of *S45* million and 580 million for the same periods in 1992. The notional amount of swaps increased to $31.5 billion from $20.8 billion and $18.4 billion at December 31, and June 39, 1992, respectively. Accruing fixed rate swaps represented $17.4 billion, $10,5 billion and $11.2 billion for the same respective periods.

Along with the second quarter report, Bane One made available to its investors a 100 page brochure entitled *Bane One Corporation Asset and Liability Management.* This brochure described how the corporation uses swaps and other derivatives to maintain its strong capital position, manage its liquidity, and manage the bank's interest rate exposure.

Panel C: 1993 Third Quarter

The following information supplements Management's Discussion and Analysis in Part 1. The notional amount of swaps shown below represents an agreed upon amount on which calculations of interest payments to be exchanged are based. BANC ONE's credit exposure is limited to the net difference between the calculated pay and receive amounts on each transaction which are generally netted and paid quarterly. BANC ONE's policy is to obtain sufficient collateral from swap counterparties to secure receipt of all amounts due, At September 30, 1993, the market value of interest rate swaps and the related collateral was approximately $5365 million and $623 million respectively. As indicated below the notional value of the interest rate swap portfolio increased from 521 billion to 538 billion during the nine months ended September 30, 1993, This increase was primarily associated with swaps acquired to replaced fixed rate, on-and off-balance sheet instruments which have or will mature or amortize and to manage interest rate risk in newly acquired affiliates. These new affiliates did not use swaps to manage their exposure to interest rate risk to as great a degree as BANC ONE. Exposure to interest rate risk is determined by simulating the impact of prospective changes in interest rates on the results of operations. Management seeks to insure that over a one-year horizon, net income will not be impacted by more than 4 percent and 9 percent by a gradual change in market interest rates of 1 percent and 2 percent, respectively. At December 31, 1992, a 2.3 percent reduction in forecasted earnings would have resulted from a gradual 1 percent increase in market rates. Due to the increase in the notional value of the swap portfolio noted above, the sensitivity to such a rate increase changed to 3.8 percent at September 3D, 1993. BANC ONE believes that both on-balance sheet securities and off-balance sheet derivatives may be used interchangeably to manage interest rate risk to an acceptable level. Various factors are considered in deciding the appropriate mix of such securities and derivatives including liquidity, capital requirements and yield,

During the nine months ended September 30, 1993, BANC ONE entered into 53.8 billion notional amount of basis swap contracts where payments based on the prime rate and LIBOR are exchanged. The variable rate used in the non-basis swap contracts entered into by BANC ONE are based on LIBOR, while many of the variable rate assets being synthetically altered are based on the prime rate. Basis swap contracts, therefore, improve the degree to which the swap portfolio acts as a hedge against the impact of changes in rates on BANC ONE's results of operations.

The table below summarizes by notional amounts the activity for each major category of swaps. For all periods presented, BANC ONE had no deferred gains or losses relating to terminated swap contracts. The terminations shown in the following table for the year ended December 31, 1991 resulted in losses of 51.8 million which were recognized during that year in accordance with BANC ONE's accounting policy at that time. The terminations in 1993 related to swaps which had been carried at market value and, therefore, resulted in no deferred gain or *loss* at termination.

$ (millions)	Received Fixed	Pay Fixed	Basis	Forward Starting	Total
Balance, December 31, 1990…………..	$ 3,114	$ 937	$550	$117	$4,719
Additions………………………………	9,797	509			10,306
Maturities / Amortizations……………	(1,171)	(322)			(1,493)
Terminations…………………………	(3,102)				(3,102)
Forward Starting-Becoming Effective….	117			117	
Acquisition and other (net)……………..	2,764	277			3,041
Balance December 31, 1991 …………	11,519	1,401	550		13,470
Additions………………………………..	2,002	501		11,656	14,159
Maturities/Amortizations……………….	(6,059)	(182)	(350)		
Terminations ……………………..					
Forward Starting-Becoming Effective…	3,201	1,005		(4,206)	
Acquisitions and Other (net) ………..	289	(296)			(7)
Balance, December 31, 1992……….	10,952	2,429	200	7,450	21,031
Additions……………………………..	4,428	1,237	3,800	12,000	21,465
Maturities/Amortizations…………….	(3,545)	(861)	(204)		(4,610)
Terminations…………………………	(250)	(250)			(500)
Forward Starting – Becoming Effective	10,480			(10,480)	
Acquisition and Other (net) …………	450	15	20	(150)	335
Balance, September 30, 1993……….	$22,515	$2,570	$3,816	$8,820	$37,721

The table below summarized expected maturities and weighted average interest rates to be received and paid on the swap portfolio at September 30, 1993: A key assumption in the preparation of the table is that rates remain constant at September 30, 1993 levels. To the extent that rates change, both the periodic maturities and the variable interest rates to be received or paid will change. Such changes could be substantial. The maturities change when interest rates change because the swap portfolio includes $23.6 billion of amortizing swaps. Amortization is generally based on certain interest rate indices.

EXPECTED MA TURlTY								
	4th Quart.							
$ (millions)	1993	1994	1995	1996	1997	1998	All Other	Total
Receive Fixed Swaps								
Notional Amount	$2,436	$9,096	$8,880	$1,050	$90	$46	$917	$22,515
Weighted Average								
Receive Rate	7.58%	6.00%	5.34%	6.02%	7.24%	6,22%	6.81%	5.95%
Pay Rate …………………	6.64	3.28%	3.23	3.36	3.24	3.19	3.54	3.19
Pay Fixed Swaps								
Notional Amount ………..	$627	$970	$318	$272	$267	$109	$7.	$2,570
Weighted Average								
Receive Rate	3.25%	3.39%	3.33%	3.26%	3.44%	3.41%	3.31%	3.34%
Pay Rate	6.64	5.86	5.00	5.76	6.07	5.30	8.82	5.96
Basis Swaps								
Notional Amount …..................	0	0	0	$2,200	$1,600	$16	0	$3,816
Weighted Average								
Receive Rate	0.00	0.00	0.00	3.22%	3.27%	3.20%	0.00	3.24%
Pay. Rate	0.00	0.00	0.00	3.33	3.34	4.80	0.00	3.34
Forward-Starting"								
Notional Amount	$500	$100	$6,720	$1,500	0	0	0	$8,820
Weighted Average								
Receive Rate ……………..	7.20%	5.74%	4.98%	5.68%	0.00	0.00	0.00	5.24%
Pay Rate. ……………………	3.38	3.38	3.38	3.38	0.00	0.00	0.00	3.38
Total...	$3.563	$10,166	$15.918	$5.022	$1,957	$171	$924	$37.721
	6.77%	5.75%	5.15%	4.54%	3.47%	4.14%	6.78%	5.33%
	3.88	3.53	3.33	3.48	3.71	4.69	3.58	3.49

In preparation for his meeting with McCoy, Dick Lodge asked his staff to prepare a simplified set of Banc One financials that could communicate the essence of the bank's financial statements and the underlying economics of their business. This stylized set of financials would show the basic earnings sensitivity faced by the bank, and how it used swaps to solve this problem. The simplified model would also demonstrate the impact of the bank's derivative activities on its accounting ratios, such as its net interest margin as well as its returns on assets and equity. Moreover, the simplified books would show how swaps affected the bank's dependence on large short term borrowings as well as demonstrate how the bank's swap portfolio affected the amount of risk-adjusted capital it held.

In order to explain the role that swaps played at Banc One, Lodge and his staff felt it might be instructive to compare Banc One with two hypothetical twin banks whose investment policies differed from its own. The first twin was like Banc One in all regards but one. This hypothetical bank brought its swaps onto the balance sheet by replacing the notional principal of the' its receive-fixed swaps with investments in fixed rate securities27 funded by variable-rate borrowings. Because Banc One's receive-fixed swaps were similar to an investment in fixed-rate securities funded by floating-rate borrowings, tDis twin would have similar interest rate exposure to Banc One. However, it would differ in its accounting performance, dependence .on large liabilities, and capital levels.

A second twin would follow yet another investment strategy. In place of Banc One's fixed-rate investments, this twin would invest in floating-rate loans and investments. In place of Banc One's swaps, it would invest in floating-rate assets financed by floating-rate deposits. The second twin more closely resembles a bank that did not manage its interest rate sensitivity.

The hope was that these simple projections would demonstrate to investors how the bank's investment activities, but especially its derivatives activities, affected its earnings sensitivity, accounting results, liquidity, and capital needs.

	Rate	Banc One Stylized	Twin A (No investment Activities)	Twin B (Swap on Balance Sheet)		Rate	Banc One (Stylized)	Twin A (No investment Activities)	Twin B (Swap on Balance Sheet)
BALANCE SHEET ($ IN BILLIONS)					• Additional Treasury securities	4.30%	0.00	0.79	0.00
Assets					Total interest income		5.47	6.26	5.66
Floating-rate assets					Interest expense from:				
• Variable-rate loans		533.8	533.8	533.8	• Retail deposits	3.27%	0.63	0.63	0.63
• Additional money					• Wholesale deposits	3.09%	0.27	0.27	0.27
market assets		0	0	31.8	• Additional wholesale deposits	3.09%	0.00	0.57	0.57
Fixed-rate Assets					• Fixed core deposits	3.57%	0.85	0.85	0.85
• Fixed-rate loans		18.6	18.6	18.6	• Large deposits	3.57%	0.08	0.08	0.08
• Fixed-rate investments		13.4	13.4	0	Total interest expense		1.83	2.40	2.40
• Additional Treasury securities		0	18.4	0	Income from Swaps (6)	2.50%	0.46	0.00	0.00
Other assets		8.4	8.4	8.4	Net interest		4.09	3.85	3.25
Total Assets		$74.2	$92.6	$92.6	Non-interest expense		2.37	2.37	2.37
NOTE: Earning Assets (1)		65.8	84.2	84.2	Taxable earnings		1.72	1.48	0.88
					Taxes	34.00%	0.59	0.50	0.30
Liabilities and Equity					Net income		1.14	0.98	0.58
Floating-rate liabilities									
• Retail deposits		19.3	19.3	19.3	PERFORMANCE MEASURES				
• Wholesale deposits (2)		8.8	8.8	8.8	Net interest margin (7)		6.22%	4.58%	3.86%
• Additional wholesale					Net interest margin				

	deposits (3)		0.0	18.4	18.4	(excluding swaps) (8)		5.52%	4.58%	3.86%
	Fixed-rate liabilities					Return on assets		1.53%	1.06%	0.63%
	• Fixed core deposits (4)		23.8	23.8	23.8	Equity/Assets (9)		8.56%	6.86%	6.86%
	• Large time deposits		2.3	2.3	2.3	Return on Equity (10)		17.89%	15.42%	9.19%
	Other liabilities		13.4	13.4	13.4	Dependence on large				
	Total liabilities		67.6	86.0	86.0	liabilities (11)		15.0%	33.5%	-5.4%
	Preferred shares		0.3	0.3	0.3	Risk-adjusted assets 02)		$63.2	$63.1	$74.7
	Common shares		6.4	6.4	6.4	Tier I capital/risk -adjusted				
	Total		$74.2	$92.6	$92.6	assets (13)		10.4%	10.5%	8.8%
						Earnings sensitivity (14)		-3.30%	-3.30%	12.88%
	OFF-BALANCE-SHEET ITEMS									
	Swaps (5)		$18.4	$0.0	$0.0	SUMMARY				
	• INCOME STATEMENT					Earnings		High	Better	Low
	Interest Income from:					Capital		High	Low	Low
	• Variable-rate loans	7.32%	$2.47	$2.47	$2.47	Risk Capital		Good	High	Low
	• Additional money					Liquidity		Good	Low	High
	market assets	3.50%	0.00	0.00	1.11	Earnings Sensitivity		Liability	Liability	Very
	• Fixed-rate loans	11.13%	2.07	2.07	2.07			Sensitive Asset	Sensitive	
	• Fixed-rate investments	6.88%	0.92	0.92	0.00					Sensitive

1. Earning assets include loans and investments.
2. "Wholesale" deposits represent liabilities to other financial institutions, eg., federal funds borrowings.
3. For both twin banks, additional needs for fund would be met by borrowing from other financial institutions.
4. Fixed core deposits are the "Sticky fixed- deposits. Their rates may change with market rates (at bank management', discretion), but they are relatively stable in volume as rates change.
5. Represents only the swaps in which Banc One receives and the current floating rate. Does not include Banc One's basis swap.
6. Represents the difference between the fixed rate that Banc One receives and the current floating rate. Does not include Banc One's basis swaps.
7. Net interest (including income from swaps) divided by earning assets.
8. Net interest (excluding income from swaps) divided by earning assets.
9. Common equity / assets.
10. Return to common equity.
11. Equals (large time deposits + wholesale deposits – money market assets) / (earning assets – money market assets). Represents an estimate of the liabilities that the bank might be called on to honour immediately, net of its assets that could be liquidated immediately.
12. Calculated by applying the BIS capital weights to each assets category.
13. Banc one's equity divided by its risk adjusted assets.
14. Represents the percentage change in the coming year's net income in response to a gradual 1% rise in interest rates over the coming year. In this model, a gradual 1% rise in rates is the same as an immediate. 5% increase in rates. The earnings sensitivity for a 1% rise. This is because of the amortization schedule of the bank's swap contract as well as the nature of the other bank assets and liabilities. Furthermore, a 1% fall in rates would not necessarily produce the same earnings sensitivity. Bank One estimated that a 1% drop in rates would lead to a 4.0% increased in earnings as compared to a 3.3% decline in earnings for a 1% rate increase.

LARGE LOSSES IN DERIVATIVES MARKETS

Derivatives Debacles: Case Studies of Large Losses in Derivatives Markets

Recent years have witnessed numerous accounts of derivatives-related losses on the part of established and reputable firms. These episodes have precipitated concern, and even alarm, over the recent rapid growth of derivatives markets and the dangers posed by the widespread use of such instruments. What lessons do these events hold for policymakers? Do they indicate the need for stricter government supervision of derivatives markets, or for new laws and regulations to limit the use of these instruments? A better understanding of the events surrounding recent derivatives debacles can help to answer such questions.

This chapter presents accounts of two of the costliest and most highly publicized derivatives-related losses to date. The episodes examined involve the firms of Metallgesellschaft AG and Barings PLC. Each account begins with a review of the events leading to the derivatives-related loss in question, followed by an analysis of the factors responsible for the debacle. Both incidents raise a number of public policy questions: Can government intervention stop such incidents from happening again? Is it appropriate for the government even to try? And if so, what reforms are indicated? These issues are addressed at the end of each case study, where the lessons and public policy concerns highlighted by each episode are discussed.

Risk and Regulation in Derivatives Markets

Perhaps the most widely cited report on the risks associated with derivatives was published in 1993 by the Group of Thirty-a group consisting of prominent members of the international financial community and noted academics. The report identified four basic kinds of risks associated with the use of derivatives.[1] Market risk is defined as the risk to earnings from adverse movements in market prices. Press accounts of derivatives-related losses have tended to emphasize market risk; but the incidents examined in this chapter illustrate the importance of operational risk-the risk of losses occurring as a result of inadequate systems and control, human error, or management failure.

Counterparty credit risk is the risk that a party to a derivative contract will fail to perform on its obligation. Exposure to counterparty credit risk is determined by the cost of replacing a contract if counterparty (as a party to a derivatives contract is known) were to default.

Legal risk is the risk of loss because a contract is found not to be legally enforceable. Derivatives are legal contracts. Like any other contract, they require a legal infrastructure to provide for the resolution of conflicts and the enforcement of contract provisions. Legal risk is a prime public policy concern, since it can interfere with the orderly functioning of markets.

These risks are not unique to derivative instruments. They are the same types of risks involved in more traditional types of financial intermediation, such as banking and securities underwriting. Legal risk does pose special problems for derivatives markets, however. The novelty of many derivatives makes them susceptible to legal risk because of the uncertainty that exists over the applicability of existing laws and regulations to such contracts.

Although the risks associated with derivatives are much the same as those in other areas of finance, there nonetheless seems to be a popular perception that the rapid growth of derivatives trading in recent years poses special problems for financial markets. Most of these concerns have centered on the growth of the over-the counter (OTC) derivatives market. As Stoll (1995) notes, concern about the growth of OTC derivatives markets has arisen because these instruments are nonstandard contracts, without secondary trading and with limited public price information. Moreover, OTC markets lack some of the financial safeguards used by futures and options exchanges, such as margining systems and the daily marking to market of contracts, designed to ensure that all market participants settle any losses promptly. The absence of such safeguards, along with the complexity of many of the new generation of financial derivatives and the sheer size of the market, has given rise to concerns that the growth of derivatives trading might somehow contribute to financial instability. Finally, there is some concern among policymakers that the federal financial regulatory agencies have failed to keep pace with the rapid innovation in OTC derivatives markets.2 Such concerns have only been reinforced by frequent reports of derivatives-related losses in recent years.

The traditional rationale for regulating financial markets stems from concerns that events in these markets can have a significant impact on the economy. Much of the present-day financial regulatory system in the United States evolved as a response to financial panics that accompanied widespread economic recessions and depressions. For example, the creation of the Federal Reserve System was prompted in large part by the Panic of 1907, while the advent of federal deposit insurance was a response to the thousands of bank failures that accompanied the Great Depression.

The present-day financial regulatory system has several goals. The most important is to maintain smoothly functioning financial markets. A prime responsibility of institutions like the Federal Reserve is to keep isolated events, such as the failure of a single bank, from disrupting the operation of financial markets generally. During the twentieth century, U.S. financial market regulation expanded to encompass at least two more goals. The creation of a system of federal deposit insurance in 1933 gave the federal government a stake in the financial condition of individual commercial banks, since a federal agency was now responsible for meeting a bank's obligations to its insured depositors in the event of insolvency. In addition, Congress enacted the Securities Exchange Act to help protect investors by requiring firms issuing publicly traded securities to provide accurate financial reports. The act created the Securities and Exchange Commission (SEC) to regulate the sales and trading practices of securities brokers, as well as to enforce the provisions of the law more generally.

Although financial market regulation deals largely with the problem of managing risk, it cannot eliminate all risk. Risk is inherent in all economic activity. Financial intermediaries such as commercial and investment banks specialize in managing financial risks. Regulation can seek to encourage such institutions to manage risks prudently, but it cannot eliminate the risks inherent in financial intermediation. There is a tension here. Regulators seek to reduce the risks taken on by the firms they regulate. At the same time, however, firms cannot earn profits without taking risks. Thus, an overzealous attempt to reduce risk could prove counterproductive-a firm will not survive if it cannot earn profits.

Conventional wisdom views derivatives markets as markets for risk transfer. According to this view, derivatives markets exist to facilitate the transfer of market risk from firms that wish to avoid such risks to others more willing or better suited to manage those risks. The important thing to note in this regard is that derivatives markets do not create new risks-they just facilitate risk management. Viewed from this perspective, the rapid growth of derivatives markets in recent years simply reflects advances in the technology of risk management. Used properly, derivatives can help organizations reduce financial risk. Although incidents involving large losses receive the most public attention, such incidents are the exception rather than the rule in derivatives markets.

Most public policy concerns center around the speculative use of derivatives. Speculation involves the voluntary assumption of market risk in the hope of realizing a financial gain. The existence of speculation need not concern policymakers as long as all speculative losses are borne privately, that is, only by those individuals or organizations that choose to engage in such activities. But many policymakers fear that large losses on the part of one firm may lead to a widespread disruption of financial markets-the collapse of Barings illustrates some of the foundations for such concerns. In the case of an insured bank, regulators discourage speculation because it can lead to losses that may ultimately become the burden of the government.[3]

A view implicit in many recent calls for more comprehensive regulation of derivatives markets is that these markets are subject only to minimal regulation at present. But exchange-traded derivatives, such as futures contracts, have long been subject to comprehensive government regulation. In the United States, the SEC regulates securities and options exchanges while the Commodity Futures Trading Commission (CFTC) regulates futures exchanges and futures brokers. Although OTC derivatives markets are not regulated by any single federal agency, most OTC dealers, such as commercial banks and brokerage firms, are subject to federal regulation.[4] As it happens, both incidents examined in this chapter involve instruments traded on regulated exchanges. Any judgment as to whether these incidents indicate a need for more comprehensive regulation of these markets requires some understanding of just what happened in each case.

Metallgesellschaft

Metallgesellschaft AG (hereafter, MG) is a large industrial conglomerate engaged in a wide range of activities, from mining and engineering to trade and financial services. In December 1993, the firm reported huge derivatives-related losses at its U.S. oil subsidiary, Metallgesellschaft Refining and Marketing (MGRM). These losses were later estimated at over $1 billion, the largest derivatives-related losses ever reported by any firm at the time. The incident helped bring MG-then Germany's fourteenth largest industrial corporation-to the brink of bankruptcy. After dismissing the firm's executive chairman, Heinz Schimmelbusch, and several other senior managers, MG's board of supervisors was forced to negotiate a $1.9 billion rescue package with the firm's 120 creditor banks (Roth 1994a, b).

MG's board blamed the firm's problems on lax operational control by senior management, charging that "speculative oil deals had plunged Metallgesellschaft into the crisis."5 Early press reports echoed this interpretation of events, but subsequent studies report that MGRM's use of energy derivatives was an integral part of a combined marketing and hedging program under which the firm offered customers long term price guarantees on deliveries of petroleum products such as gasoline and heating oil. Reports that MG's losses were attributable to a hedging program have raised a host of new questions. Many analysts remain puzzled by the question of how a firm could lose over $1" billion by hedging.

The Metallgesellschaft debacle has sparked a lively debate on the shortcomings of the firm's hedging strategy and the lessons to be learned from the incident. The ensuing account draws from a number of recent articles, notably Culp and Hanke (1994); Culp and Miller 0994a, b, 1995a, b, c, d); Edwards and Canter 0995a, b); and Mello and Parsons (1995a, b).

MGRM's Marketing Program

In 1992, MGRM began implementing an aggressive marketing program in which it offered long-term price guarantees on deliveries of gasoline, heating oil, and diesel fuels for up to five or ten years. This program included several novel contracts, two of which are relevant to this study. The first was a "firm-fixed" program, under which a customer agreed to fixed monthly deliveries at fixed prices. The second, known as the "firm-flexible" contract, specified a fixed price and total volume of future deliveries but gave the customer some flexibility to set the delivery schedule. Under the second program, a customer could request 20 percent of its contracted volume for anyone year with 45 days' notice. By September 1993, MGRM had committed to sell forward the equivalent of over 150 million barrels of oil for delivery at fixed prices, with most contracts for terms of ten years.

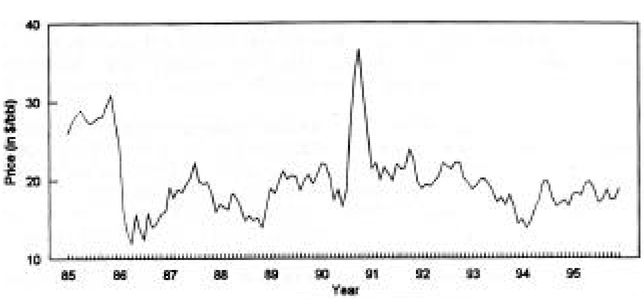
Both types of contracts included options for early termination. These cash-out provisions permitted customers to call for cash settlement on the full volume of outstanding deliveries if market prices for oil rose above the contracted price. The firm-fixed contract permitted a customer to receive one-half the difference between the current nearby futures price (that is, the price of the futures contract closest to expiration) and the contracted delivery price, multiplied by the entire remaining quantity of scheduled deliveries. The firm-flexible contract permitted a customer to receive the full difference between the second-nearest futures price and the contract price. multiplied by all remaining deliverable quantities.[6]

MGRM negotiated most of its contracts in the summer of 1993. Its contracted delivery prices reflected a premium of $3 to $5 per barrel over the prevailing spot price of oil. As is evident in Figure 23.1, energy prices were relatively low by recent historical standards during this period and were continuing to fall. As long as oil prices kept falling, or at least did not rise appreciably, MGRM stood to make a handsome profit from this marketing arrangement. But a significant increase in energy prices could have exposed the firm to massive losses unless it hedged its exposure.

MGRM sought to offset the exposure resulting from its delivery commitments by buying a combination of short-dated oil swaps and futures contracts as part of a strategy known as a "stack-and-roll" hedge. In its simplest form, a stack-and-roll hedge involves repeatedly buying a bundle, or "stack," of short-dated futures or forward contracts to hedge a longer term exposure. Each stack is "rolled over" just before expiration by selling the existing contracts while buying another stack of contracts for a more distant delivery date; hence the term stack-and-roll. MGRM implemented its hedging strategy by maintaining long positions in a wide variety of contract months, which it shifted between contracts for different oil products (crude oil, gasoline, and heating oil) in a manner intended to minimize the costs of rolling over its positions.

Had oil prices risen, the accompanying gain in the value of MGRM's hedge would have produced positive cash flows that would have offset losses stemming from its commitments to deliver oil at below-market prices. As it happened, however, oil prices fell eyen further in late 1993. Moreover, declines in spot and near-term oil futures and forward prices significantly exceeded declines in long-term forward prices. As a result contemporaneous realized losses from the hedge appeared to exceed any potential offsetting gains accruing to MGRM's long-term forward commitments.

Figure 23.1 Crude Oil Prices (1985 – 1995)



This precipitous decline in oil prices caused funding problems for MGRM. The practice in futures markets of marking futures contracts to market at the end of each trading session forced the firm to recognize its futures trading losses immediately, triggering huge margin calls. Normally, forward contracts have the advantage of permitting hedgers to defer recognition of losses on long-term commitments. But MGRM's stack-and-roll hedge substituted short-term forward contracts (in the form of short term energy swaps maturing in late 1993) for long-term forward contracts. As these

contracts matured, MGRM was forced to make large payments to its counterparties, putting further pressure on its cash flows. At the same time, most offsetting gains on its forward delivery commitments were deferred.

Rumors of MGRM's problems began to surface in early December. In response to these developments, the New York Mercantile Exchange (NYMEX). the exchange on which MGRM had been trading energy futures, raised its margin requirements for the firm. This action, which was intended to protect the exchange in case of a default, further exacerbated MGRM's funding problems. Rumors of the firm's financial difficulties led many of its GTC counterparties to begin terminating their contracts. Others began demanding that it post collateral to secure contract performance.

Upon learning of these circumstances, MG's board of supervisors fired the firm's chief executive and installed new management. The board instructed MG's new managers to begin liquidating MGRM's hedge and to enter into negotiations to cancel its long-term contracts with its customers. This action further complicated matters, however. NYMEX withdrew its hedging exemption once MGR_1 announced the end of its hedging program. Hedging exemptions permit firms to take on much larger positions in exchange-traded futures than those allowed for unhedged speculative positions. The loss of its hedging exemption forced MGRM to reduce its positions in energy futures still further (Culp and Miller, 1994a).

The actions taken by MG's board of supervisors have spurred widespread debate and criticism, as well as several lawsuits. Some analysts argue that MGRM's hedging program was seriously flawed and that MG's board was right to terminate it. Others, including Nobel Prize-winning economist Merton Miller, argue that the hedging program was sound and that MG's board exacerbated any hedging-related losses by terminating the program prematurely.

The Debate over MGRM's Hedging Program

As Figure 23.1 shows, oil prices began rising in 1994, soon after _IGRM's new management lifted the firm's hedge. It thus appears that MGRM could have recouped most if not all of its losses simply by sticking to its hedging program. Whether management should have been able to anticipate this outcome is the topic of an active debate, however.

Disagreements over the efficacy of MGRM's hedging program stem from differing assumptions about the goal of the hedging program (or, perhaps more accurately, what the goal should have been), and the feasibility of continuing the program in light of the large negative cash flows MGRM experienced in late 1993.

Culp and Miller (1994a, b, 1995a, b, c, d) and Culp and Hanke (1994) are critical of MG's board of supervisors for terminating MGRM's marketing and hedging program. To be sure, Culp and Miller do find that MGRM's hedging program had suffered losses, albeit much smaller losses than those calculated by MG's auditors. But they argue that those losses did not justify terminating MGRM's hedging program. According to Culp and Miller, most of MG's reported losses were attributable to the manner in which its new management chose to terminate its subsidiary's marketing program, not to defects in its hedging strategy. It is not unusual for the parties to such agreements to negotiate termination of a contract before it expires. The normal practice in such circumstances involves payment by one party to the other to compensate for any changes in the value of the contract. In contrast, it appears that MGRM's new management simply agreed to terminate its contracts with its customers without asking for any payment to reflect the increased value of those contracts. The hedge-however imperfect-effectively was transformed by this action into a huge speculative transaction after the fact.

Edwards and Canter (1995a, b) and Mello and Parsons (1995a, b) are more critical of MGRM's hedging strategy. These writers emphasize the difficulties that MGRM's large negative cash flows created for the parent company. They argue that MGRM's management could have-and should have-sought to avoid such difficulties by designing a hedge that would have minimized the volatility of its cash flows.

Although they are critical of MGRM's hedging strategy, Edwards and Canter offer no opinion as to whether MG's board was right to terminate the program. Like Culp and Miller, they are puzzled about the decision to terminate existing contracts with customers without negotiating some payment to compensate for the increase in the value of those contracts.[7]

Mello and Parsons' criticisms of MGRM's hedging strategy are unequivocal. They argue that MGRM's strategy was fatally flawed, and they defend the decision to terminate the hedging program as the only means of limiting even greater potential future losses. They also emphasize the difficulty that MG's new management would have had in securing the financing necessary to maintain MGRM's hedging program and argue that funding considerations should have led the subsidiary's managers to synthesize a hedge using long-dated forward contracts. In this context, Mello and Parsons note that the parent firm already had accumulated a cash flow deficit of OM 5.65 billion between 1988 and 1993. This deficit had been financed largely by bank loans. Considering these circumstances, they find the reluctance of MG's creditor banks to fund the continued operation of the oil marketing program understandable.

Reconciling Opposing Views

These disagreements over the efficacy of MGRM's hedging strategy seem unlikely ever to be resolved, based as they are on different assumptions about the goals management should have had for its strategy. The main issue, then, is whether MG's senior management and board of supervisors fully appreciated the risks the firm's U.S. oil subsidiary had assumed. If they did, the firm should have arranged for a line of credit to fund its short-term cash flows. Indeed, Culp and Miller (1995a, c, d) claim that MGRM had secured lines of credit with its banks just to prepare for such contingencies. Yet the subsequent behavior of MG's board suggests that its members had very little prior knowledge of MGRM's marketing program and were uncomfortable with its hedging strategy, despite the existence of a written strategic plan.

It is difficult for an outside observer to assign responsibility for any misunderstandings between MG's managers and its board of supervisors. MG's board ultimately held Heinz Schimmelbusch, the firm's executive chairman, responsible for the firm's losses, claiming that he and other senior managers had lost control over the activities of the firm and concealed evidence of losses.8 In response, Schimmelbusch has filed suit against Ronaldo Schmitz and Deutsche Bank, seeking $10 million in general and punitive damages (Taylor, 1995b). Arthur Benson, former head of MGRM and architect of the firm's ill-fated hedging program, is suing MG's board for $1 billion on charges of defamation (Taylor, 1994). Thus, the issue of blame appears destined to be settled by the U.S. courts.

Response of the CFTC

The Metallgesellschaft debacle did not escape the attention of U.S. regulators. In July 1995, the u.S. Commodity Futures Trading Commission instituted administrative proceedings against MGRM and MG Futures, Inc. (MGFI), an affiliated Futures Commission Merchant that processed trades for MGRM and other MG subsidiaries.[9] The CFTC order charged both MGRM and MGFI with "material inadequacies in internal control systems" associated with MGRM's activity in energy and futures markets. In addition, MGFI was charged with failing to inform the CFTC of these material inadequacies, while MGRM was charged with selling illegal, off-exchange futures contracts. The two MG subsidiaries settled the CFTC action without admitting or denying the charges and agreed to pay the CFTC a $2.5 million settlement. They also agreed to implement a series of CFTC recommendations to reform their internal controls and to refrain from violating CFTC regulations. The CFTC's action rendered MGRM's firm-fixed agreements illegal and void.[10] Thus, the CFTC's action would have created legal risk for Metallgesellschaft and its customers except that the firm had already canceled most of the contracts in question.

The CFTC's actions in this case have proven somewhat controversial. Under the Commodity Exchange Act, the CFTC is charged with regulating exchange-traded futures contracts. At the same time, the act explicitly excludes ordinary commercial forward contracts from the jurisdiction of the CFTc. The legal definition of a futures contract is open to differing interpretations, however, leading to some uncertainty over the legal status of OTC derivatives under the Commodity Exchange Act. Most market participants felt that this uncertainty was resolved in 1993 when, at the behest of Congress, the CFTC agreed to exempt off-exchange forward and swaps contracts from regulations governing exchange-traded contracts. CFTC chairman Mary Schapiro maintains that the agency's action against MGRM does not represent a reversal of its policy on OTC contracts. According to Schapiro, the CFTC's order is worded narrowly so as to apply only to contracts such as the firm fixed (45-day) agreements sold by MGRM in this case.[11]

Nonetheless, this action has prompted some critics to charge the agency with creating uncertainty about the legal status of commercial forward contracts. Critics of the action include Miller and Culp (1995) and Wendy Gramm, a former chairman of the CFTC.[12] The CFTC's action has also been criticized by at least two prominent members of Congress-Rep. Thomas J. Bliley, Jr., Chairman of the House Commerce Committee; and Rep. Pat Roberts, Chairman of the House Agricultural Committee.[13]

Since the CFTC's action against Metallgesellschaft is narrowly directed and involves somewhat esoteric legal arguments, it is too soon to know what its effect will be on OTC derivatives markets generally. Still, commodity dealers must now take extra care in designing long-term delivery contracts to avoid potential legal problems.[14]

An Overview of Policy Concerns

Considering the debate over the merits of MGRM's hedging strategy, it would seem naive simply to blame the firm's problems on its speculative use of derivatives. It is true that MGRM's hedging program was not without risks. But the firm's losses are attributable more to operational risk-the risk of loss caused by inadequate systems and control or management failure-than to market risk. If MG's supervisory board is to be believed, the firm's previous management lost control of the firm and then acted to conceal its losses from board members. If one sides with the firm's previous managers (as well as with Culp, Hanke, and Miller), then the supervisory board and its bankers misjudged the risks associated with MGRM's hedging program and panicked when faced with large, short-term funding demands. Either way, the loss was attributable to poor management.

Does this episode indicate the need for new government policies or more comprehensive regulation of derivatives markets? The answer appears to be no. MGRM's losses do not appear ever to have threatened the stability of financial markets. Moreover, those losses were due in large part to the firm's use of futures contracts which trade in a market that is already subject to comprehensive regulation. The actions taken by the CFTC in this instance demonstrate clearly that U.S. regulators already have the authority to intervene when they deem it necessary. Unfortunately, the nature of those actions in this case may create added legal risk for other market participants.

To view the entire incident in its proper perspective, it must be remembered that MG's losses were incurred in connection with a marketing program aimed at providing long-term, fixed-price delivery contracts to customers, a type of arrangement common to many types of commercial activity. Systematic attempts to discourage such arrangements would seem to be poor public policy.

Finally, MG's financial difficulties were nor attributable solely to its use of derivatives. As noted earlier, the firm's troubles stemmed in part from the heavy debt load it had accumulated in previous years. Moreover. MGRM's oil marketing program was not the only source of its parent company's losses during 1993. MG reported losses of DM 1.8 billion on its operations for the fiscal year ended September 30, 1993, in addition to the DM 1.5 billion loss auditors attributed to its hedging program as of the same date (Roth, 1994b). Simply stated, the MG debacle resulted from poor management. As a practical matter, government policy cannot prevent firms such as Metallgesellschaft from making mistakes. Nor should it attempt to do so.

Barings

At the time of its demise in February 1995, Barings PLC was the oldest merchant bank in Great Britain. Founded in 1762 by the sons of German immigrants, the bank had a long and distinguished history. Barings had helped a fledgling United States of America to arrange the financing of the Louisiana Purchase in 1803. It had also helped Britain finance the Napoleonic Wars, a feat that prompted the British government to bestow five noble titles on the Baring family.

Although it was once the largest merchant bank in Britain, Barings was no longer the powerhouse it had been in the nineteenth century. With total shareholder equity of £440 million, it was far from the largest or most important banking organization in Great Britain. Nonetheless, it continued to rank among the nation's most prestigious institutions. Its clients included the Queen of England and other members of the royal family.

Barings had long enjoyed a reputation as a conservatively run institution. But that reputation was shattered on February 24, 1995, when Peter Baring, the bank's chairman, contacted the Bank of England to explain that a trader in the firm's Singapore futures subsidiary had lost huge sums of money speculating on Nikkei-225 stock index futures and options. In the days that followed, investigators found that the bank's total losses exceeded US$1 billion, a sum large enough to bankrupt the institution.

Barings had almost failed once before in 1890 after losing millions in loans to Argentina, but it was rescued on that occasion by a consortium led by the Bank of England. A similar effort was mounted in February 1995, but the attempt failed when no immediate buyer could be found and the Bank of England refused to assume liability for Barings' losses. On the evening of Sunday, February 26, the Bank of England took action to place Barings into administration, a legal proceeding resembling Chapter 11 bankruptcy-court proceedings in the United States. The crisis brought about by Barings' insolvency ended just over one week later when a large Dutch financial conglomerate, the International Nederland Grope (ING), assumed the assets and liabilities of the failed merchant bank.

What has shocked most observers is that such a highly regarded institution could fall victim to such a fate. The ensuing account examines the events leading up to the failure of Barings, the factors responsible for the debacle, and the repercussions of that event on world financial markets.[15] This account is followed by an examination of the policy concerns arising from the episode and the lessons these events hold for market participants and policy makers.

Unauthorized Trading Activities

In 1992, Barings sent Nicholas Leeson, a clerk from its London office, to manage the back-office accounting and settlement operations at its Singapore futures subsidiary. Baring Futures (Singapore), hereafter BFS, was established to enable Barings to execute trades on the Singapore International Monetary Exchange (SIMEX). The subsidiary's profits were expected to come primarily from brokerage commissions for trades executed on behalf of customers and other Barings subsidiaries.16

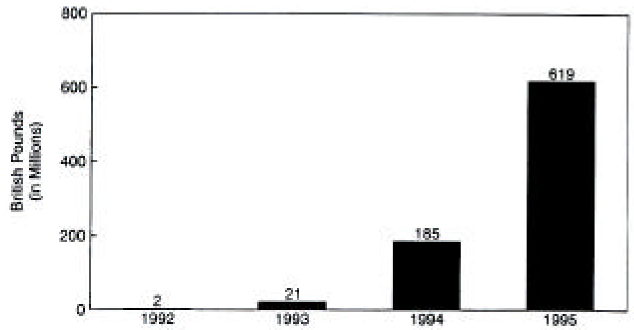
Soon after arriving in Singapore, Leeson asked permission to take the SIMEX examinations that would permit him to trade on the floor of the exchange. He passed the examinations and began trading later that year. Some time during late 1992 or early 1993, Leeson was named general manager and head trader of BFS. Normally the functions of trading and settlements are kept separate within an organization, as the head of settlements is expected to provide independent verification of records of trading activity. But Leeson was never relieved of his authority over the subsidiary's back-office operations when his responsibilities were expanded to include trading.

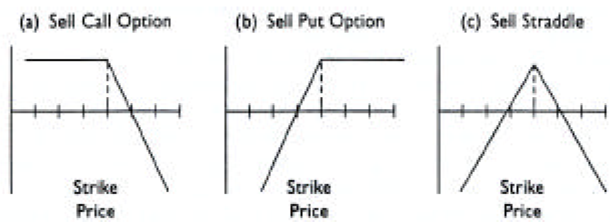
Leeson soon began to engage in proprietary trading-that is, trading for the firm's own account. Barings' management understood that such trading involved arbitrage in Nikkei-225 stock index futures and ten-year Japanese Government Bond (JGB) futures. Both contracts trade on SIMEX and the Osaka Securities Exchange (OSE). At times price discrepancies can develop between the same contract on different exchanges, leaving room for an arbitrageur to earn profits by buying the lower-priced contract on one exchange while selling the higher-priced contract on the other. In theory this type of arbitrage involves only perfectly hedged positions, and so it is commonly regarded as a low-risk activity. Unbeknownst to the bank's management, however, Leeson soon embarked upon a much riskier trading strategy. Rather than engaging in arbitrage, as Barings' management believed, he began placing bets on the direction of price movements on the Tokyo stock exchange.

Leeson's reported trading profits were spectacular. His earnings soon came to account for a significant share of Barings' total profits; the bank's senior management regarded him as a star performer. After Barings failed, however, investigators found that Leeson's reported profits had been fictitious from the start. Because his duties included supervision of both trading and settlements for the Singapore subsidiary, Leeson was able to manufacture fictitious reports concerning his trading activities. He had set up a special account-account number 88888-in July 1992, and instructed his clerks to omit information on that account from their reports to the London head office. By manipulating information on his trading activity, Leeson was able to conceal his trading losses and report large profits instead.

Figure 23.2 shows Leeson's trading losses from 1992 through the end of February 1995. By the end of 1992-just a few months after he had begun trading-Leeson had accumulated a hidden loss of £2 million. That figure remained unchanged until October 1993, when his losses began to rise sharply. He lost another £21 million in 1993 and £185 million in 1994. Total cumulative losses at the end of 1994 stood at £208 million. That amount was slightly larger than the £205 million profit reported by the Barings Group as a whole before accounting for taxes and for £102 million in scheduled bonuses.

A major part of Leeson's trading strategy involved the sale of options on Nikkei225 futures contracts. Figure 23.3a and 23.3b show the payoff at expiration accruing to the seller of a call or put option, respectively. The seller of an option earns a premium in return for accepting the obligation to buy or sell the underlying item at a stipulated strike price. If the option expires out-of-the-money, the option premium becomes the seller's profit. If prices turn out to be more volatile than expected, however, an option seller's potential losses are virtually unlimited.





Some time in 1994, Leeson began selling large number of option straddles, a strategy that involved the simultaneous sale of both calls and puts on Nikkei-225 futures. Figure 23.3c shows the payoff at expiration to a sold option straddle. Option prices reflect the market's expectation of the price volatility of the underlying item. The seller of an option straddle earns a profit only if the market proves less volatile than predicted by option prices. As is evident in Figure 23.3c, Leeson's strategy amounted to a bet that the Japanese stock market would neither fall nor increase by a great deal-any large movement in Japanese stock prices would result in losses. By January 1, 1995, Leeson was short 37,925 Nikkei calls and 32,967 Nikkei puts. He also held a long position of just over 1,000 contracts in Nikkei stock index futures, which would gain in value if the stock market were to rise.

Disaster struck on January 17 when news of a violent earthquake in Kobe, Japan, sent the Japanese stock market into a tailspin. Over the next five days, the Nikkei index fell over 1,500 points-Leeson's options positions sustained a loss of £68 million. As stock prices fell, he began buying massive amounts of Nikkei stock index futures. He also placed a side bet on Japanese interest rates, selling Japanese government bond futures by the thousands in the expectation of rising interest rates.

This strategy seemed to work for a short time. By February 6, the Japanese stock market had recovered by over 1,000 points, making it possible for Leeson to recoup most of the losses resulting from the market's reaction to the earthquake. His cumulative losses on that date totaled £253 million, about 20 percent higher than they had been at the start of the year. But within days the market began falling again-Leeson's losses began to multiply. He continued to increase his exposure as the market kept falling. By February 23, Leeson had bought over 61,000 Nikkei futures contracts, representing 49 percent of total open interest in the March 1995 Nikkei futures contract and 24 percent of the open interest in the June contract. His position in Japanese government bond futures totaled just over 26,000 contracts sold, representing 88 percent of the open interest in the June 1995 contract. Leeson also took on positions in Euroyen futures. He began 1995 with long positions in Euroyen contracts (a bet that Japanese interest rates would fall) but then switched to selling the contracts. By February 23 he had accumulated a short position in Euroyen futures equivalent to 5 percent of the open interest in the June 1995 contract and 1 percent of the open interest in both the September and December contracts.

Barings faced massive margin calls as Leeson's losses mounted. While these margin calls raised eyebrows at the bank's London and Tokyo offices, they did not prompt an immediate inquiry into Leeson's activities. It was not until February 6 that Barings' group treasurer, Tony Hawes, flew to Singapore to investigate irregularities with the accounts at BFS. Accompanying Hawes was Tony Railton, a settlements clerk from the London office.

While in Singapore, Hawes met with SIMEX officials, who had expressed concern over Barings' extraordinarily large positions. Hawes assured them that his firm was aware of these positions and stood ready to meet its obligations to the exchange. His assurances were predicated on the belief that the firm's exposure on the Singapore exchange had been hedged with offsetting positions on the Osaka exchange. He was soon to learn that this belief was incorrect.

Leeson's requests for additional funding continued during February, and Barings' London office continued to meet those requests-in all, Barings committed a total of £742 million to finance margin calls for BFS. Meanwhile Tony Railton, the clerk Hawes had dispatched to Singapore, found that he could not reconcile the accounts of BFS. Particularly disturbing was a US$190 million discrepancy in one of BFS' accounts. For over a week, Railton attempted to meet with Leeson to resolve these discrepancies. Leeson had become hard to find, however. Railton finally tracked him down on the floor of the Singapore exchange on Thursday, February 23, and persuaded Leeson to meet with him that evening. When the meeting began, Railton began asking a series of difficult questions. At that point, Leeson excused himself, stating that he would return shortly. But he never did return. Instead, he and his wife left Singapore that evening. The next day, Leeson faxed his resignation to Barings' London office from a hotel in Kuala Lumpur, stating in part, "My sincere apologies for the predicament I have left you in. It was neither my intention nor aim for this to happen." [17]

After Leeson failed to return, Railton and others at Barings' Singapore office began investigating his private records and quickly discovered evidence that he had lost astronomical sums of money. Peter Baring, the bank's chairman, did not learn of the bank's difficulties until the next day, when he was forced to call the Bank of England to ask for assistance. Ironically, this was the same day that Barings was to inform its staff of their bonuses. Leeson was to receive a £450,000 bonus, up from £130,000 the previous year, on the strength of his reported profits. Baring himself expected to receive £1 million.

The Bank of England's Board of Banking Supervision (1995) subsequently conducted an inquiry into the collapse of Barings. According to the Board's report, total losses attributable to Leeson's actions came 10£927 million (approximately US$1.4 billion) including liquidation costs, an amount far in excess of Barings' total equity of £440 million. Most of the cost of the Barings' debacle was borne by its shareholders and by ING, the firm that bought Barings. Barings was a privately held firm; most of its equity was held by the Baring Foundation, a charity registered in the United Kingdom. Barings' executive committee held the firm's voting shares, which constituted a small fraction of the firm's total equity. Although ING was able to buy the failed merchant bank for a token amount of £1, it had to pay £660 million to recapitalize the firm. SIMEX subsequently reported that the funds Barings had on deposit with the exchange were sufficient to meet the costs incurred in liquidating its positions (Szala, Nusbaum, and Reerink, 1975). It is not known whether the OSE suffered any losses as a result of Barings' collapse.

Leeson was later detained by authorities at the airport in Frankfort, Germany, and was extradited to Singapore the following November. In Singapore, Leeson pleaded guilty to charges of fraud and was sentenced to a 6h-year prison term (Mark 1995).

Certain material facts regarding the entire incident are not yet known, as Leeson refused to cooperate with British authorities unless extradited to Great Britain. He later contested the findings of the Banking Board's inquiry, however. A letter to the board from his solicitor's states,

These conclusions are inaccurate in various respects. Indeed, in relation to certain of the matters they betray a fundamental misunderstanding of the actual events. Unfortunately, given the uncertainty regarding Mr. Leeson's position we are not able to provide you with a detailed response to your letter.[18]

Leeson has promised to write a book describing his own version of events while serving out his prison term in Singapore.

Market Aftershocks

Once the Singapore and Osaka exchanges learned that Barings would not be able to meet its margin calls, they took control of all the bank's open positions. The Nikkei index fell precipitously when market participants learned that the exchanges would be liquidating such large positions. Thus, in the days immediately following the announcement of Barings' collapse, it was not known whether the margin money the bank had deposited with the exchanges would cover the losses stemming from the liquidation of its positions.

Matters were further complicated when SIMEX announced it would double margin requirements on its Nikkei stock index futures contract effective Tuesday, February 28. Fearing that their margin money might be used to pay for Barings' losses, several of the exchange's U.S. clearing members threatened to withhold payment of the additional margin SIMEX was demanding of them unless given assurances that such margin payments would be used solely to collateralize their own accounts. A refusal to pay would have caused the affected dealers to forfeit their positions. If that had happened, SIMEX would have been faced with a series of defaults. According to CFTC chairman Schapiro, such an event could have "destroyed the ability of SIMEX to manage the situation."19 Indeed, there are reports that many market participants feared that the very solvency of the SIMEX clearinghouse was in question. To complicate matters further, Japanese and Singaporean regulators were slow to inform market participants of the steps they were taking to insure the financial integrity of the exchange clearinghouses. This lack of communication served only to exacerbate the fears of market participants (Falloon, 1995; Irving, 1995; McGee, 1995a, b; Szala, Nusbaum, and Reerink, 1995).

Upon learning of the situation, Chairman Schapiro contacted the Monetary Authority of Singapore (MAS) to persuade the agency to assure SIMEX's clearing members that their margin deposits would not be used to offset Barings' proprietary losses. The MAS subsequently acceded to these requests and provided its assurance in a short statement released before the start of trading on Thesday. SIMEX's margin calls were met and a potential crisis was avoided.

This was not the end of headaches for Barings' customers, however. BFS was one of the largest clearing member firms on SIMEX. As such, it handled clearing and settlement for[16] U.S. firms and held approximately $480 million in margin funds on their behalf when it went bankrupt.

U.S. futures exchanges typically arrange the immediate transfer to other firms of all customer accounts of a financially troubled clearing member. Laws in the United States facilitate such transfers because they provide for strict segregation of customer accounts, which prevents the creditors of a broker or clearing member firm from attaching the assets of customers. That Japanese law contains no such provisions was not well known before the collapse of Barings. Although laws in Singapore do recognize the segregation of accounts, SIMEX had never before dealt with the insolvency of a clearing member firm. To complicate matters further, most of BFS' customer accounts had been booked through Baring Securities in London. Consequently, SIMEX did not have detailed information on individual customer positions. It had records only on a single commingled account for Baring Securities. Finally, much of the information that Leeson had provided to the exchange, as well as to Barings' other offices, was false. These circumstances made the task of sorting out the positions of individual customers extremely difficult.

During the next week, Barings' U.S. customers scrambled to reproduce documentation of their transactions with the bank and supplied this information to SIMEX and the OSE. But while this information made it possible for the exchanges to identify customer positions, Barings' bankruptcy administrator in London had asked the exchanges to block access to all Barings' margin deposits. The bankruptcy administrator had raised questions about whether U.K. laws on the segregation of customer accounts were applicable in an insolvency of this kind (Szala, Nusbaum, and Reerink, 1995).

It was not until ING took over Barings, on March 9, that the bank's customers were assured of access to their funds. Even then, access was delayed in many cases. By one account, several major clients waited more than three weeks before their funds were returned (Irving, 1995).

Policy Concerns Highlighted by Barings' Default

All futures exchanges maintain systems to prevent the accumulation of large speculative losses. But events surrounding the collapse of Barings have served to highlight weaknesses in risk management on the part of SIMEX and other futures exchanges. They also suggest a need for closer international cooperation among futures exchanges and their regulators and for clearer laws on the status of customer accounts when a clearing member firm becomes insolvent.

Futures exchanges maintain stringent speculative position limits for individual firms and traders to prevent large losses and to limit their exposure. It appears that SIMEX relaxed some of these restrictions for BFS, however. It is not unusual for futures exchanges to grant exemptions to established position limits for hedged positions, such as those Leeson claimed to maintain. But it is normal for the exchange clearinghouse to monitor closely the activities of firms receiving such exemptions and to take steps to verify the existence of offsetting exposures. It now appears that SIMEX failed to pursue such precautions in its dealings with Barings.

The exchange's attitude toward Barings was influenced in part by the bank's strong international reputation, but its willingness to relax normal risk management guidelines also may have been attributable to its desire to attract business. Although the OSE was first to list Japanese government bond and Nikkei-225 stock index futures, SIMEX soon began listing similar contracts in direct competition with the Osaka exchange. Thereafter, the two exchanges battled each other for market share. Barings was one of the most active firms on SIMEX-and Leeson was responsible for much of the exchange's trading volume in Nikkei stock index futures and options. Thus, some observers believe that SIMEX may have been too willing to accommodate BFS (McGee, 1995a). Critics include representatives of U.S. futures exchanges, who maintain that their risk management standards are more stringent.[20] A report on the incident commissioned by the government of Singapore came to a similar conclusion, finding that the exchange may have been too liberal in granting increases in position limits.[21]

Communication between exchanges can be important for identifying and resolving potential problems. Communication between SIMEX and the OSE was minimal, however. This lack of communication not only helped make it possible for Leeson to accumulate large losses but also hampered efforts to contain the damage once Barings collapsed. Although the OSE routinely published a list of the positions of its most active traders, SIMEX did not make such disclosures. It now seems apparent that SIMEX officials never consulted the OSE's list to verify Leeson's claim that he was hedging his large positions in Singapore with offsetting exposures on the Osaka exchange.

Some observers blame this lack of communication on the rivalry between the two exchanges. Arrangements existing between u.s. exchanges suggest that competition need not preclude information sharing, however. In the United States, futures exchanges attempt to coordinate their activities with the CFTC and other futures exchanges. Each exchange maintains strict speculative position limits established under CFTC oversight. The CFTC monitors compliance through a comprehensive surveillance policy that includes a large-trader reporting system. Market participants are required to justify unusually large positions. This system enabled the CFTC to ascertain quickly that Barings had no significant positions on any U.S. futures exchange at the time of its collapse.[22]

While competitive concerns may sometimes give exchanges incentives to relax prudential standards, as many observers seem to think that SIMEX did, it does not follow that regulators should seek to discourage such competition. Competition among exchanges serves an important economic function by encouraging innovation. Securities and futures exchanges constantly compete with one another to provide new products to their customers. Thus, whereas futures exchanges once listed contracts only for agricultural and other commodities, a significant fraction of all futures trading today involves contracts for financial instruments. The growth of trading in such instruments has provided important benefits to international financial markets, helping to make them more efficient while facilitating risk management by financial intermediaries and commercial firms alike. Moreover, competition giyes futures exchanges an incentive to maintain strong financial controls and risk management systems, as most market participants seek to avoid risks like those faced by SIMEX customers after the collapse of Barings. Finally, policymakers need not restrict competition to address the problems highlighted by the Barings debacle.

The events surrounding the collapse of Barings led futures industry regulators from 16 nations to meet in Windsor, England, in May 1995 to discuss the need for legal and regulatory reform. At that meeting, officials agreed on a plan of action now known as the Windsor Declaration. The declaration calls for regulators to promote, as appropriate, "national provisions and market procedures that facilitate the prompt liquidation and/or transfer of positions, funds and assets, from failing members of futures exchanges," and to support measures "to enhance emergency procedures at financial intermediaries, market members and markets and to improve existing mechanisms for international co-operation and communication among market authorities and regulators.,"[23] The International Organization of Securities Commissions (IOSCO) later endorsed the Windsor Declaration and pledged to study the issues it raised. IOSCO also asked its members to promote declaration measures in cross-border transactions.[24]

The Barings debacle has also spurred efforts by market participants to strengthen financial safeguards at futures and options exchanges. In March 1995, the Futures Industry Association (FIA) organized a task force to investigate measures to improve the financial integrity of futures and options exchanges. The association's Global Task Force on Financial Integrity (1995) subsequently published a report containing 60 recommendations, ranging from risk management practices to customer protection issues. The FIA report encourages all nations to review their bankruptcy laws to clarify the status of customer funds and to modify provisions that might conflict with the laws of other nations. It recommends that exchanges and their regulators establish procedures for the transfer of a troubled clearing member firm's customer assets before it is declared insolvent, as is now typically done in the United States. In addition, the report encourages exchange clearinghouses to monitor their clearing member firms closely and to perform periodic audits. Thus, the FIA's recommendations are broadly consistent with the principles espoused by the Windsor Declaration, especially in their emphasis on customer protection and the need for improved information sharing among exchanges and government authorities.

Subsequently, the clearing organizations for 19 U.S. stock, stock option, and futures exchanges announced their intent to begin pooling data on transactions of member firms (McGee, 1995c). In addition, CFTC Chairman Schapiro has announced that her staff will work with the futures industry to develop concrete customer protection proposals.[25]

The Barings debacle has served to galvanize an international effort-one that has been joined by government officials and market participants alike-to re-evaluate risk management systems, customer protection laws, and procedures for dealing with the failure of a large clearinghouse member. It also has prompted increased communication and pledges of greater cooperation among regulators from different nations. It is still too early to pass judgment on the ultimate success of such initiatives, however. While regulators have pledged increased international cooperation, recent press accounts have noted that officials in Britain, Japan, and Singapore have not always cooperated with one another in conducting their investigations of the Barings case.[26]

Lessons from the Barings Debacle

The losses suffered by Barings provide a good example of the market risk associated with derivatives. But, as with the case of Metallgesellschaft, the Barings debacle best illustrates operational risk and legal risk. In this regard, the Bank of England's Board of Banking Supervision inquiry concluded,

> Barings' collapse was due to the unauthorized and ultimately catastrophic activities of, it appears, one individual (Leeson) that went undetected as a consequence of a failure of management and other internal controls of the most basic kind. Management failed at various levels and in a variety of ways to institute a proper system of internal controls, to enforce accountability for all profits, risks and operations, and adequately to follow up on a number of warning signals over a prolonged period. [27]

The board's inquiry found nine separate warning signs that should have alerted Barings' management to problems with its Singapore futures subsidiary. A partial list of those warning signs includes the following:

- The lack of segregation of duties between front and back offices. This lack was identified as a weakness and potential problem area in an internal audit report following a review of BFS' operations in the summer of 1994. Barings' management failed to act on the report's recommendations to remedy this situation.

- The high level of funding requested by Leeson. Between December 31, 1994, and February 24, 1995, Barings provided Leeson with £521 million to meet margin calls. Total funding of BFS stood at £742 million, more than twice the reported capital of the Barings Group, when Leeson's activities were finally discovered on February 24.[28]
- The unreconciled balance of funds transferred to BFS to meet margin calls. In his requests for additional funding, Leeson often claimed the money was needed for client accounts but never provided detailed information about these accounts as was the usual practice. Nonetheless, the bank's head office in London paid those funds without any independent check on the validity of Leeson's requests and with no attempt to reconcile those requests with known trading positions. Perhaps the most troubling aspect of Barings' behavior in this regard is that SIMEX rules prohibit its members from financing the margin accounts of customers. Barings' management apparently ignored evidence that the firm might be doing so in violation of SIMEX rules.
- The apparent high profitability of Leeson's trading activities relative to the low level of risk as perceived and authorized by Barings' management in London. High returns typically entail high risk. Yet no one in senior management seriously questioned how Leeson's strong reported profits could result from what was supposed to have been a low-risk activity. To be sure, at least one executive observed that "This guy must be busting his intraday limits or something. "29 But Leeson's reports were never challenged until too late, and management did little to restrain his trading activities. According to interviews with Barings' staff, Leeson was regarded as "almost a miracle worker," and there was "a concern not to do anything which might upset him. "30
- The discovery of discrepancies in Leeson's accounts by outside auditors. Barings' auditors, the firm of Coopers & Lybrand, informed the bank's management of a £50 million discrepancy in BFS's accounts on or before February 1, 1995. Although this discrepancy ultimately did prompt Barings' treasurer to investigate Leeson's accounts, the Board of Banking Supervision concluded that management was too slow in responding to this warning sign.
- Communications from SIMEX. The rapid buildup of Leeson's positions during January 1995 prompted SIMEX to seek assurances from Barings' management in London regarding the ability of BFS to fund its margin caUs. In retrospect, it appears that Barings' management was too hasty in providing such assurances.
- Market rumors and concerns made known to Barings' management in January and February. By late January, rumors were circulating on the OSE regarding Barings' large positions in Nikkei futures. On January 27, the Bank for International Settlements in Basle, Switzerland, raised a high-level inquiry with Barings executives in London regarding a rumor that the bank had experienced losses and could not meet its margin calls on the 05E. On the same day, another Barings executive received a call from the Bloomberg Information Service inquiring into the bank's large positions on the OSE.

Taken together, these warning signs suggest that Barings' management had ample cause to be concerned about Leeson's activities. But management was too slow to act on these warning signs. An on-site examination of Leeson's accounts came too late to save the bank.

The Board of Banking Supervision's report outlined a number _f lessons to' be learned from the failure of Barings. They emphasize five lessons for the management of financial institutions:

- Management teams have a duty to understand fully the businesses they manage;
- Responsibility for each business activity has to be clearly established and communicated;
- Clear segregation of duties is fundamental to any effective control system;
- Relevant internal controls, including independent risk management, have to be established for all business activities;
- Top management and the Audit Committee have to ensure that significant weaknesses, identified to them by internal audit or otherwise, are resolved quickly.[31]

The report also had some criticisms for the Bank of England's supervision of Barings. U.K. banking regulations require all banks to notify the Bank of England before entering into a transaction that would expose more than 25 percent of the organization's capital to the risk of loss. A Bank of England manager granted Barings an informal concession permitting it to exceed this limit in its exposure to SIMEX and the OSE without first referring the matter to the Bank's senior management. But while the report is somewhat critical of the Bank of England on this matter, it concludes,

The events leading up to the collapse of Barings do not, in our view, of themselves point to the need for any fundamental change in the framework of regulation in the UK. There is, however, a need for improvements in the existing arrangements.[32]

The report goes on to suggest a number of ways to improve the Bank of England's supervision of banks. According to the report,

- the Bank should "explore ways of increasing its understanding of the nonbanking businesses. . . undertaken by those banks for which it is responsible";[33]
- it should prepare explicit internal guidelines to assist its supervisory staff in identifying activities that could pose material risks to banks and ensure that adequate safeguards are in place;
- it should work more closely with the Securities and Futures Authority, the agency responsible for regulating the domestic operations of British-based securities firms, as well as with regulators from other nations; and

- it should address deficiencies in the implementation of rules dealing with large exposures.

The report also recommended an independent quality assurance review of the Bank of England's supervisory function.

The Board 01 Banking Supervision's report did not blame the collapse of Barings on its use of derivatives. Instead, it placed responsibility for the debacle on poor operational controls at Barings.

The failings at Barings were not a consequence of the complexity of the business, but were primarily a failure on the part of a number of individuals to do their jobs properly. . . . While the use of futures and options contracts id enable Leeson to take much greater levels of risk (through their leverage) than might have been the case in some other markets, it was his ability to act without authority and without detection that brought Barings down. [34]

This point has been reinforced recently by news of a similar debacle at the New York office of Daiwa Bank, where a trader concealed large trading losses for over ten years before finally confessing to his activities.[35] Parallels between the Daiwa and Barings debacles are striking, as both incidents resulted from the unauthorized activities of a single trader. Daiwa's losses were in no way related to derivatives, however. The bank incurred over $1 billion in losses as a result of unauthorized trading in U.S. government bonds, widely regarded as the safest of financial instruments.

Some Final Observations on the Barings Debacle

The events surrounding the collapse of Barings have highlighted certain weaknesses in international financial markets that represent legitimate concerns for policymakers. Two of these weaknesses deserve special notice: (1) the lack of communication between securities and futures exchanges and regulators in different countries; and (2) conflicting laws on the legal status of customer accounts at futures brokers and clearing agents in the event of insolvency. These weaknesses can be addressed only by increased international cooperation among futures exchanges, regulators, and lawmakers.

At the same time, it does not appear that more stringent government regulation of futures markets could have prevented the Barings debacle. Leeson acted outside existing regulatory guidelines and outside the law in concealing the true nature of his trading activities and the losses resulting therefrom. Existing laws and regulations should have been able to prevent, or at least to detect, Leeson's activities before he could incur such astronomical losses. But Barings, SIMEX, and the Bank of England were all lax in enforcing those rules. Barings was lax in enforcing basic operational controls. In doing so, it violated not only official regulations but also commonly accepted market standards for managing risk. Similarly, it appears that SIMEX may have been too liberal in granting increases in position limits to BFS. Finally, the Bank of England granted Barings an exemption that helped make it possible for Leeson to continue his illicit activities undetected.

Concluding Comments

The cases of Metallgesellschaft and Barings provide an interesting study in contrasts. Both cases involve exchange-traded derivatives contracts. In both cases, senior management has been criticized for making an insufficient effort to understand fully the activities of their firms' subsidiaries and for failing to monitor and supervise the activities of those subsidiaries adequately. But while critics have faulted MG's management for overreacting to the large margin calls faced by one of its subsidiaries, Barings' management has been faulted for being overly complacent in the face of a large number of warning signs.

If these two disparate incidents offer any single lesson, it is the need for senior management to understand the nature of the firm's activities and the risks that those activities involve. In the case of Metallgesellschaft, the sheer scale of its U.S. oil subsidiary's marketing program exposed the firm to large risks. Although there is a great deal of disagreement over the efficacy of the hedging strategy employed by MGRM, it would seem difficult to argue that members of MG's board of supervisors fully appreciated the nature or magnitude of the risks assumed by the firm's U.S. oil subsidiary. If they had, they would not have been so shocked to find the firm facing large margin calls. In the case of Barings, senior management seemed content to accept that a single trader could earn huge profits without exposing the firm to large risks. With the benefit of hindsight, it seems clear that senior executives of both firms should have taken more effort to understand the activities of subordinates.

News of derivatives-related losses often prompts calls for more comprehensive regulation of derivatives markets. Bur the cases of Metallgesellschaft and Barings - which rank among the largest derivatives-related losses to date-involve instruments traded in markets already subject to comprehensive regulation. In the case of Barings, the debacle involved a regulated merchant bank trading in regulated futures markets. If anything, the Barings debacle illustrates the limits of regulation. Established rules and regulations should have been able to prevent a single trader from accumulating catastrophic losses. But both SIMEX and the Bank of England granted exemptions that helped make it possible for Leeson to continue his activities for years without being detected. It appears that regulatory organizations can also be subject to operational weaknesses.

Moreover, the instruments traded by these two firms-oil futures, stock index futures, and stock index options-are not the kinds of complex and exotic instruments responsible for concerns often expressed in connection with the growth of derivatives markets. In the case of Barings, the Bank of England's Board of Banking Supervision concluded that it was not the complexity of the business but the failure of a large number of individuals to do their jobs properly that made the bank susceptible to catastrophic losses by a single trader. As the recent misfortune of Daiwa Bank shows, weaknesses in operational controls can lead to losses in many areas of a firm's operations, not just those involved with derivatives. The losses suffered by Daiwa resulted from trading in U.S. Treasury bonds, widely regarded as the safest of all securities.

Unfortunately, no amount of regulation can remove all risk from financial markets. Risk is inherent in all economic activity, and financial markets exist to help market participants diversify such risks. At the same time, regulation can impose costs on market participants. The Metallgesellschaft case shows that attempts at stringent regulation can sometimes have undesirable side effects. According to critics, the CFTC's action against MG's U.S. subsidiaries has introduced uncertainty about the legal status of commercial forward contracts. As a general rule, government policy should attempt to minimize legal risk rather than create it.

To be sure, the Barings debacle did highlight the need for certain legal and regulatory reforms and for more international cooperation among exchanges and their regulators. But market discipline is also a powerful form of regulation. Highly publicized accounts of derivatives-related losses have led many firms to scrutinize their risk management practices-not only in the area of derivatives, but in other areas of their operations as well. Thus, while it is true that derivatives debacles often reveal the existence of disturbing operational weaknesses among the firms involved, such incidents can also teach lessons that help to make financial markets safer in the long run. As the foregoing accounts show, regulation cannot substitute for sound management practices. At the same time, government policymakers can act to minimize the potential for disruption to financial markets by promoting laws and policies that minimize legal risk.

References

BNA's Banking Repor "CFTC Chairman Schapiro Tells Congress Barings-Type Disaster Unlikely in U.S.," 64 (March 6, 1995a), pp. 468-69.

- "CFTC Undertaking Regulatory Review: Schapiro Address [es] Failure of Barings PLC," 64 (March 6, 1995b), pp. 469-70.
- "World Futures Regulators Adopt Plan of Action in View of Barings," 64 (May 22, 1995c), pp. 1017-18.
- "Schapiro Says CFTC, Industry to Craft Customer Protection Proposals," 65 (July 3, 1995d), pp. 15-16.
- "IOSCO Endorses Value-at-Risk Models for Capital Adequacy Calculations," 65 (July 17, 1995e), pp. 120-21.
- "Schapiro Defends CFTC Action Against Two MG Subsidiaries," 65 (September 25, 19950, p. 503.

Board of Banking Supervision. Report of the Board of Banking Supervision Inquiry into the Circumstances of the Collapse of Barings. London: HMSO, 1995.

Culp, Christopher, and Steve H. Hanke. "Derivative Dingbats," The International Economy (July/August 1994).

Culp, Christopher L., and Merton H. Miller. "Hedging a Flow of Commodity Deliveries with Futures: Lessons From Metallgesellschaft," Derivatives Quarterly, 1 (Fall 1994a), pp. 7-15.

- "Letter to the Editor," Risk, 11 (November 1994b), p. 18.
- "Metallgesellschaft and the Economics of Synthetic Storage," journal of Applied Corporate Finance, 7 (Winter 1995a), pp. 62-76.
- "Auditing the Auditors," Risk, 8 (April 1995b), pp. 36-9.
- "Hedging in the Theory of Corporate Finance: A Reply to our Critics." journal of Applied Corporate Finance, 8 (Spring 1995c), pp. 121-27.
- "Basis Risk and Hedging Strategies: Reply to Mello and Parsons," Derivatives Quarterly, 1 (Summer 1995d), pp. 20-26.
- The Economist. "Revolution at Metallgesellschaft," December 25, 1993, p. 90.
- "The Barings Collapse: Spot the Smoking Receivable," October 21, 1995.p.79.

Edwards, Franklin R. "Derivatives Can Be Hazardous to Your Health: The Case of Metallgesellschaft." Derivatives Quarterly, 1 (Spring 1995), pp. 8-17.

Edward, Franklin R., and Michael S. Canter. "The Collapse of Metallgesellschaft: Unhedgeable Risks, Poor Hedging Strategy, or Just Bad Luck?" The journal of Futures Markets, 15 (May 1995a), pp. 211-64.

Edward, Franklin R., and Michael S. Canter. "The Collapse of Metallgesellschaft: Unhedge able Risks, Poor Hedging Strategy, or Just Bad Luck? journal of Applied Corporate Finance, 8 (Spring 1995b), pp. 86-105.

Falloon, William. "Who's Missing from the Picture?" Risk, 8 (April 1995), pp. 19-22. Fox, Justin. "2 in House Fault Futures Trading Panel on Over-the-Counter Derivatives Ruling," American Banker, December 26,1995.

Futures Industry Association GJobal Task Force on Financial Integrity. Financial Integrity Recommendations. Washington: Futures Industry Association, 1995. Global Derivatives Study Group. Derivatives: Practices and Principles. Washington:

Group of Thirty, 1993.

Irving, Richard. "Beyond Barings," Risk, 8 (April 1995), p. 6.

Kawaller, Ira. "Hedging with Futures Contracts: Going the Extra Mile," journal of Cash Management, 6 (July/ August 1986), pp. 34-36.

Mark, Jeremy. "With Leeson in Singapore Prison, Focus Shifts to Barings Executives," Wall Street journal, December 4, 1995.

McGee, Suzanne. "Reform Sought to Prevent Future Barings," Wall Street journal, March 17, 1995a.

- "New U.S. Futures Chief Is Crafting a Global Role," Wall Street journal, May 17, 1995b.
- "Nineteen U.S. Exchanges To Share Data on Dealings," Wall Street journal, September 6, 1995c.

Mello, Antonio S., and John E. Parsons. "Maturity Structure of a Hedge Matters: Lessons from the Metallgesellschaft Debacle," journal of Applied Corporate Finance, 8 (Spring 1995a), pp. 106-20.

- "Hedging a Flow of Commodity Deliveries with Futures: Problems with a Rolling Stack," Derivatives Quarterly, 1 (Summer 1995b), pp. 16-19.

Miller, Merton H., and Christopher L. Culp. "Rein in the CFTC," Wall Street journal, August 17,1995.

Rance, Brian D. "The Commodity Futures Trading Commission Order of Settlement with MG Refining and Marketing, Inc. and MF Futures, Inc.," Derivatives Quarterly, 2 (Winter 1995), pp. 13-17.

Roth, Terrence. "Metallgesellschaft Sets Shareholder Vote Despite Worries Over Financial Package," Wall Street journal, January 14, 1994a.

- "German Firm's Bailout Package Gets Approved," Wall Street journal, January 17, 1994b.

Sapsford, Jathon, Michael R. Sesit, and Timothy L. O'Brien. "How Daiwa Bond Man in New York Cost Bank $1.1 Billion in Losses," Wall Street journal, September 27, 1995.

Springett, Pauline. "The Barings Rescue: 'Apologies, but the pressures have become too much to bear,'" The Guardian (London), March 7, 1995.

Stoll, Hans R. "Lost Barings: A Tale in Three Parts Concluding with a Lesson," The journal of Derivatives, 3 (Fall 1995), pp. 109-115.

Szala, Ginger, David Nusbaum, and Jack Reerink. "Barings Abyss," Futures, 24 (May1995), pp. 68-74.

Taylor, Jeffrey. "MG Corp. Aims to Stay Active in Oil Business," Wall Street journal, March 14, 1994.

- "Securities Firms Agree to Set Controls on Derivatives." Wall Street journal, March 9, 1995a.
- "Documents at MG Appear to Contradict Fired Chairman's Stance on Derivatives," Wall Street journal, January 27, 1995b.

U.S. Commodity Futures Trading Commission. "Order Instituting Proceedings Pursuant to

Sections 6(c) and 8(a) of the Commodity Exchange Act and Findings and Order Imposing Remedial Sanctions," CFTC Docket No. 95-14 (1995a).

- "CFTC Order Imposes $2.25 Million Civil Penalty against MG Refining and Marketing, Inc. and MG Futures, Inc.; Other Remedial Sanctions Include Review of Internal Control Systems for Risk Management," Release No. 3859-95 (July 27, 1995b).

U.S. General Accounting Office. Financial Derivatives: Actions Needed to Protect the Financial

System. GAO/GCD-94-133, May, 1994. Wall Street journal. "Metallgesellschaft A.G. Dismisses Chairman and Finance Chief," December 20, 1993.

- "Former CFTC Chief Scores Agency's Move in Derivatives Case," December1, 1995.

Endnotes

1. See Global Derivatives Study Group (1993).
2. See U.S. General Accounting Office (1994).
3. Recent losses by firms such as Gibson Greetings and Procter & Gamble have also raised concerns about sales practices and the disclosure of risks associated with complex financial derivatives. Neither of the cases examined in this study involves such concerns, however.
4. Many securities companies book their OTC derivatives through unregulated subsidiaries. Although these subsidiaries are not subject to formal SEC regulation. the largest brokerage firms have agreed to abide by certain regulatory guidelines and to make regular disclosures to both the SEC and CFTC about their management of derivatives-related risks. See Taylor (1995a).
5. As cited in Edwards and Canter (1995b), page 86.
6. Mello and Parsons (1995a) provide a detailed description of these contracts.
7. More recently, Edwards (1995) has become more critical of the decision to liquidate MGRM's forward delivery contracts.
8. See the Wall Street journal (1993) and The Economist (1993).
9. A Futures Commission Merchant is a broker that accepts and executes orders for transactions on futures exchanges for customers. Futures Commission Merchants are regulated by the CFTc.
10. See U.S. Commodity Futures Trading Commission 0995a, b).
11. See BNA's Banking Report (1995f).
12. For a summary of Gramm's comments see the Wall Street journal (1995).
13. See Fox (1995).
14. See Rance (1995) for a legal analysis of these issues.
15. This account is based on the findings of a report by the Board of Banking Supervision of the Bank of England (1995) and on a number of press accounts dealing with the episode. Except where otherwise noted, all information on this episode was taken from the Board of Banking Supervision's published inquiry.
16. Most of BFS' business was concentrated in executing trades for a limited number of financial futures and options contracts. These were the Nikkei-225 contract, the 10-year Japanese Government Bond (JGB) contract, the three-month Euroyen contract. and options on those contracts (known as futures options). The Nikkei-225 contract is a futures contract whose value is based on the Nikkei-225 stock index. an index of the aggregate value of the stocks of 225 of the largest corporations in Japan. The JGB contract is for the future delivery of ten-year Japanese government bonds. The Euroyen contract is a futures contract whose value is determined by changes in the three-month Euroyen deposit rate. A futures option is a contract that gives the buyer the right, but not the obligation. to buy or sell a futures contract at a stipulated price on or before some specified expiration date.
17. The full text of leeson's letter of resignation can be found in Springett (1995).
18. Board of Banking Supervision (1995), Para. 1.77.
19. As cited in McGee (1995b).
20. See BNA's Banking Report (1995a), and Falloon (1995).
21. See The Economist (1995).
22. See the summary of Chairman Schapiro's testimony before Congress in BSA's Banking Report (1995a, b).
23. As cited in BNA's Banking Report (1995c).
24. See BNA's Banking Report (1995e).

25. See BNA's Banking Report (1995d).
26. See The Economist (1995).
27. Board of Banking Supervision (1995), Para. 14.1.
28. Board of Banking Supervision (1995), Para. 6.21.
29. Board of Banking Supervision (1995), Para. 3.57.
30. Board of Banking Supervision (1995), Para. 7.12.
31. Board of Banking Supervision (1995), Para. 14.2.
32. Board of Banking Supervision (1995), Para. 14.5.
33. Board of Banking Supervision (1995), Para. 14.3;
34. Board of Banking Supervision (1995), Para. 14.
35. 3; See Sapsford, Sesit, and O'Brien (1995) for early details of the Daiwa debacle.

Notes -

TUTORIAL

UNIT-IX

CONNECTION BETWEEN SWAP BONDS AND FRA

Objectives

- After completion of this lesson you will be able to devote fully to the connections between swap, bonds, and FRA markets.

Dear friends! This lesson extends the discussion of swap type instruments and outlines a simple framework for fixed-income security pricing. Term structure modeling is treated within this framework. The chapter also introduces the recent models that are becoming a benchmark in this sector.

Until recently, short-rate modeling was the most common approach in pricing and risk-managing fixed-income securities. The publication in 1992 of the Heath-Jarrow-Merton (HJM) approach enabled arbitrage-free modeling of multifactor-driven term structure models, but markets continued to use short-rate modeling. Today the situation is changing. The Forward Libor or Brace-Gatarek -Musiela (BGM) model is becoming the market standard for pricing and risk management.

This chapter will approach the issues from a practical point of view using swap markets and swap derivatives as a background. We are interested in providing a framework for analyzing the mechanics of swaps and swap derivatives, for decomposing them into simpler instruments, and for constructing synthetics. Recent models of fixed income modeling can then be built on this foundation very naturally.

It is worth starting with a review of the basic principles of swap engineering laid out in Chapter 5. First of all, swaps arc almost always designed such that their value at initiation is zero. This is a characteristic of modern swap-type "spread instruments," and there is no surprise here. Second, what makes the value of the swap equal to zero is a spread or an interest rate that is chosen with the purpose that the initial value of the swap vanishes. Third, swaps encompass more than one settlement date. This means that whatever the value of the swap rate or swap spread, these will in the end be some Sort of "average of shorter term floating rates or spreads" This not only imposes simple arbitrage conditions on relevant market rates, but also provides an opportunity to trade the volatility associated with such averages through the use of options on swaps. Since swaps are very liquid, they form an excellent underlying for swaptions. Swaptions, in turn, are related to interest rate volatilities for the underlying subperiods, which will relate to cap/floor volatilities. This structure is conducive to designing and understanding more complex swap products such as constant maturity swaps (CMS). The CMS swap is used as an example, for showing the advantages of the Forward Libor Model.

Finally, the chapter will further use the developed framework to illustrate the advantages of measure change technology, Switching between various T -forward measures, we show how convexity effects can be calculated,

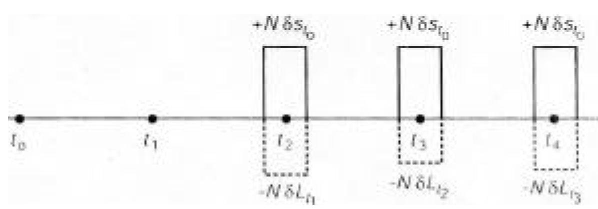
Most of the discussion will center on a three-period swap first, and then generalize the results. We begin with this simple example, because with a small number of cash flows the analysis becomes more manageable and easier to understand. Next, we lay out a somewhat more technical framework for engineering fixed-income instruments. Eventually, this is developed into the Forward Libor Model. Within our framework, measure changes using Girsanov-type transformations emerge as fundamental tools of financial engineering. The chapter discusses how measures can be changed sequentially during a numerical pricing exercise as was done in the simulation of the Forward Libor Model. These tools are then applied to CMS swaps, which are difficult to price with traditional models.

A Framework for Swaps

We work with forward fixed-payer interest rate swaps and their "spot" equivalent. These are vanilla products in the sense that contracts are predesigned and homogeneous. They are liquid, the bid-ask spreads are tight, and every market player is familiar with their properties and related conventions.

To simplify the discussion we work with a three-period swap, shown in Figure 13-1. It is worth repeating the relevant parameters again, given the somewhat more technical approach the chapter will adopt.

1. The notional amount is N, and the tenor of the underlying Libor rate is δ which represents a proportion of a calendar year. As usual, if a year is denoted by I, then δ will be 1/4 in the case of 3-month Libor.
2. The swap maturity is three periods. The swap ends at time $T = t_4$. The swap contract is signed at time t_0 but starts at time t_1, hence the term forward swap is used. [1]

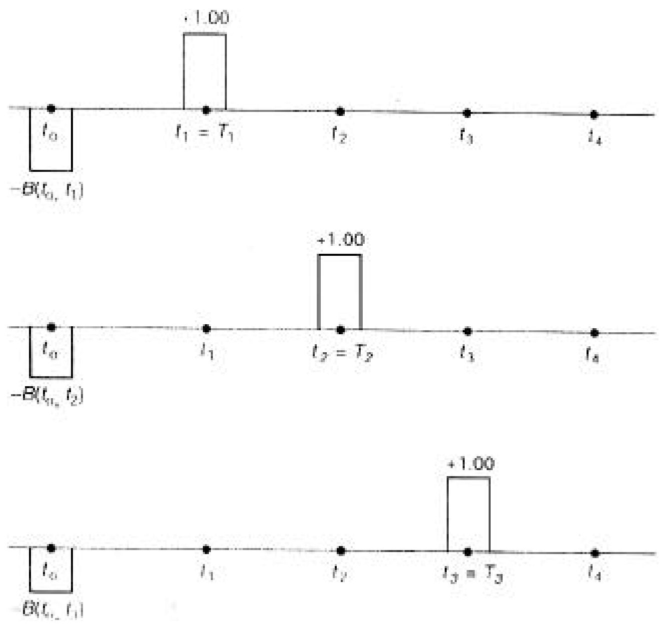


3. The dates $\{t_1, t_2, t_3\}$ are reset dates where the relevant Libor rates L_{t1}, L_{t2}, and L_{t3} will be determined.[2] These dates are b time units apart.
4. The dates $\{t_2, t_3, t_4\}$ are settlement dates where the Libor rates L_{t1}, L_t, and L_{t3} are used to exchange the floating cash flows, $N L_t$, against the fixed δN_{Sto} at each $t_i + 1$. In this setup, the time that passes until the start of the swap, $t_1 - t_0$, need not equal δ. However, it may be notationally convenient to assume that it does.

Our purpose is to provide a systematic framework in which the risk management and pricing of such swaps and the instruments that build on them can be done efficiently. That is, we discuss a technical framework that can be used for running a swap and swap derivatives book.

Swaps are one major component of a general framework for fixed-income engineering. We need two additional tools. These we introduce using a simple example again. Consider Figure 13-2, where we show payoff diagrams for three default-free pure discount bonds. The current price, $B(t_o, T_i)$, of these bonds is paid at t_0 to receive 1 dollar in the same currency at maturity dates $T_i = t_i$. Given that these bonds are default-free, the time-t_i payoffs are certain and the price $B(t_o.T_i)$ can be considered as the value today of 1 dollar to be received at time t_i. This means they are, in fact, the relevant discount factors, or in market language, simply discounts for t_i. Note that as

$$T_1 < T_2 < T_3 < T_4, \quad (1)$$



bond prices must satisfy, regardless of the slope of the yield curve:[3]

$$B(t_0, T_1) > B(t_0,T_2 > B(t_0,T_3) > B(t_0,T_4) \quad (2)$$

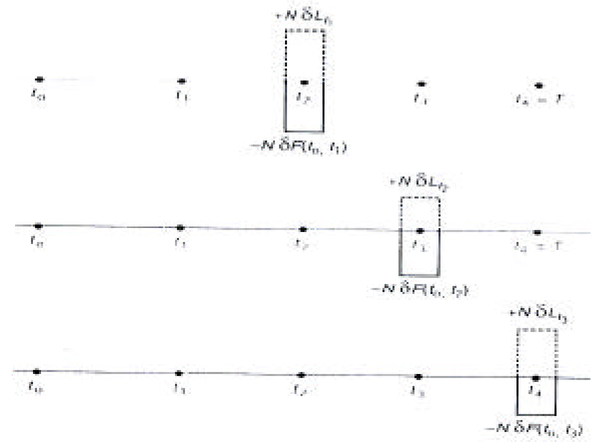
These prices can be used as discount factors to calculate present values of various cash flows occurring at future settlement dates t_i. They are, therefore, quite useful in successive swap settlements and form the second component in our framework.

The third component of the fixed-income framework is shown in Figure 13-3. Here, we have the cash flow diagrams of three forward rate agreements (FRAs) paid in arrears. The FRAs are, respectively, $t_1 \times t_2$, $t_2 \times t_3$, and $t_3 \times t_4$. For each FRA, a floating (random) payment is made against a known (fixed) payment for a net cash flow of

$$[Lt, - F(t_0,t_i)]\ N \quad (3)$$

at time $t_i + 1$. Here, the $F(t_0, t_i)$ is the forward rate of a fictitious forward loan contract signed at time t_0. The forward loan comes into effect at t_i and will be paid back at time $t_i + 1 = t_i + \delta$. We note that the fixed payments $N\,\delta\,F(t_0,t_i)$ are not the same across the FRAs. Although all FRA rates are known at time t_0, they will, in general, not equal each other or equal the payment of the fixed swap leg, d N_{sto}.

We can now use this framework to develop some important results and then apply them in financial engineering.



2.1. Equivalence of Cash Flows

The first financial engineering rule that we discuss in this chapter is associated with the perceived equivalence of cash flows. In Figure 13-3, there is a strip of floating cash flows:

$$\{N\,\delta L_{t1}, N\,\delta L_{t2}, N\,\delta L_{t3}\} \quad (4)$$

and, given observed liquid prices, the market is willing to exchange these random cash flows against the known (fixed) cash flows:

$$\{ N\,\delta\,F(t_0,t_1),\ N\,\delta\,F(t_0,t_2), N\,\delta\,F(t_0,t_3)\} \quad (5)$$

According to this, if these FRAs are liquid at time t_0, the known cash flow sequence in (5) is perceived by the markets as the correct exchange against the unknown, floating payments in (4). If we then consider the swap cash flows shown in Figure 13-1, we notice that exactly the same floating cash flow sequence as in (4) is exchanged for the known and fixed swap leg

$$\{N\delta_{Sto}, N\delta_{Sto}, N\delta_{Sto}\} \quad (6)$$

The settlement dates arc the same as well. In both exchanges, neither party marks any upfront payments at time t_0. We can therefore combine the two exchanges at time t_0, and obtain the following result.

The market is willing to exchange the fixed and known cash flows

$$\{N\delta_{Sto}, N\delta_{Sto}, N\delta_{Sto}\} \quad (7)$$

against the variable known cash flows:

$$\{N\,\delta\,F(t_0,t_1), N\,\delta\,F(t_0,t_2), N\,\delta\,F(t_0,t_3)\} \quad (8)$$

at no additional time-t_0 compensation.

This has an important implication. It means that the time-t_0 values of the two cash flow sequences arc the same. Otherwise, one party would demand an initial cash payment. Given that the cash Bows are known as of time to, their equivalence provides an equation that can be used in pricing, as we will see next. This argument will be discussed further using the forward Libor model.

2.2. Pricing the Swap

We have determined two known cash flow sequences the market is willing to exchange at no additional cost. Using this information, we now calculate the time-t_0 values of the two cash flows. To do this, we use the second component of our framework, namely, the discount bond prices given in Figure 13-2.

Suppose the pure discount bonds with arbitrage-free prices $B(t_0,t_i) = 1,2,3,4$ are liquid and actively traded. We can then use $\{B(t_0. t_2), B(t_0, t_3), B(t_0, t_4)\}$ to value cash flows settled at times t_2, t_3 and t_4 respectively.[4] In fact, the time-t_0 value or the sequence of each flows,

$$\{N\delta F(t_0,t_1), N\delta F(t_0,t_2)\ N\delta F(t_0,t_3)\} \qquad (9)$$

is given by multiplying each cash flow by the discount factor that corresponds to that particular settlement date and then adding. We use the default-free bond prices as our discount factor, and obtain the value of the fixed FRA cash flows

$$B(t_0,t_2)\ N\delta F(t_0,t_1)+B(t_0,t_3)\ N\delta F(t_0,t_2)+B(t_0,t_4)N\delta F(t_0,t_3)=$$
$$[B(t_0,t_2)\ F(t_0,t_1) + B(t_0,t_3)\ F(t_0,t_2) + B(t_0,t_4)\ F(t_0,t_3)]N \qquad (10)$$

The time-t_0 value of the fixed swap cash flows can be calculated similarly

$$B(t_0,t_2)\ N\ \delta_{Sto} + B(t_0,t_3)\ N\ \delta_{Sto} + B(t_0,t_4)\ \delta\ N_{Sto} =$$
$$[B(t_0,t_2) + B(t_0,t_3) + B(t_0,t_4)]\ \delta\ N_{Sto} \qquad (11)$$

Now, according to the argument in the previous section, the values of the two cash flows must be the same.

$$[B(t_0,t_2)F(t_0,t_1) + B(t_0,t_3)F(t_0,t_2) + B(t_0,t_4)\ F(t_0,t_3)]N =$$
$$[B(t_0,t_2) + B(t_0,t_3) + B(t_0,t_4)]\ \delta\ N_{Sto} \qquad (12)$$

This equality has at least two important implications. First, it implies that the value of the swap at time t_o is zero. Second, note that equality can be used as an equation to determine the value of one unknown. As a matter of fact, pricing the swap means determining a value for S_{to} such that the equation is satisfied. Taking S_{to} as the unknown we can rearrange Equation (12), simplify, and obtain

$$S_{to} = \frac{B(t_0,t_2)F(t_0,t_1) + B(t_0,t_3)F(t_0,t_2) + B(t_0,t_4)F(t_0,t_3)}{B(t_0,t_2) + B(t_0,t_3) + B(t_0,t_4)} \qquad (13)$$

This pricing formula can easily be generalized by moving. from the three-reriod setting to a vanilla (forward) swap that makes n payments starting at lime t_2. We obtain

$$S_{to} = \frac{\sum^n_{i=1} B(t_0,t_{i+1})\ F(t_0,t_i)}{\sum^n_{i=1} B(t_0,t_{i+1})} \qquad (14)$$

This is a compact formula that ties together the three important components of the fixed-income framework we are using in this chapter.

2.2.1. Interpretation of the Swap Rate

The formula that gives the arbitrage-free value of the (forward) swap has a nice interpretation. For simplicity revert to the three-period case. Rewrite Equation (13) as

$$S_{to} = \frac{B(t_0,t_2)}{[B(t_0,t_2)+ B(t_0,t_3) +B(t_0,t_4)]}\ F(t_0,t_1)$$

$$+ \frac{B(t_0,t_3)}{[B(t_0,t_2) + B(t_0,t_3) + B(t_0,t_4)]}\ F(t_0,t_2) \qquad (15)$$

$$+ \frac{B(t_0,t_4)}{[B(t_0,t_2) + B(t_0,t_3) + B(t_0,t_4)]}\ F(t_0,t_3) \qquad (16)$$

According to this expression, we see that the "correct" (forward) swap rate is a weighted average of the FRA paid-in-arrears rates during the life of the swap:

$$S_{to} = ?_1F(t_0,t_1) + ?_2F(t_0,t_2) + ?_3F(t_0,t_3) \qquad (17)$$

The weights are given by

$$?_{i=} \frac{B(t_0\ t_{i+1})}{[B(t_0,t_2) + B(t_0,t_3) + B(t_0, t_4)]} \qquad (18)$$

and add up to one:

$$?_1 + ?_2 + ?_3 = 1 \qquad (19)$$

This can again be generalized for a (forward) swap that makes n payments:

$$\sum_{i=1}^{n} w_i\ F(t_0 t_i) \qquad (20)$$

with $$\sum_{i=1}^{n} w_i = 1 \qquad (21)$$

Thus, the (forward) swap rate is an average paid-in-arrears FRA rate. We emphasize that this is true as long as the FRAs under consideration are paid-in-arrears. There are, on the other hand, so-called Libor-in-arrcars FRAs where a convexity adjustment needs to be made for the argument to hold.[5]

II is important to realize that the weights $\{\ _i\}$ are obtained from pure discount bond prices, which, as shown in Chapters 4 and 12, are themselves functions of forward rates:

$$B(t_0,t_i)\ \frac{1}{II^{i-1}\ (1+\delta F(t_0,t_j))} \qquad (22)$$

According to these formulas, three important components of our pricing framework-the swap market, the FRA market, and the bond market - are interlinked through nonlinear functions of forward rates. The important role played by the forward rates in these formulas suggests that obtaining arbitrage-free dynamics of these latter is required for the pricing of all swap and swap-related derivatives. The Forward Libor Model docs exactly this. Because this model is set up in a way as to fit market conventions, it is also practical.

However, before we discuss these more advanced concepts, it is best to look at an example. In practice, swap and FRA markets are liquid and market makers readily quote the relevant rates. The real-world equivalents of the pure discount bonds $\{B(t_0, t_i)\}$, on the other hand, are not that liquid, even when they exist.[6] In the following example, we sidestep this point and assume that such quotes are available at all desired maturities.

Even then, some important technical issues emerge, as the example illustrates.

Example

Suppose we observe the following paid-in-arrears FRA quotes:

Term	Bid-Ask
0 x 6	4.05- 4.07
6 x 12	4.15 – 4.17
12 x 18	4.32 – 4.34
18 x 24	4.50 - 4.54

Also, suppose the following treasury strip prices are observed:

Maturity	Bid-Ask
12 months	96.00-96.02
18 months	93.96 93.99
24 months	91.88-91.92

We can ask two questions. First, are these data arbitrage-free so that they can be used in obtaining an arbitrage-free swap rate? Second, if they are, what is the implied forward swap rate for the period that starts in six months and ends in 24 months?

The answer to the first question can be checked by using the following arbitrage equality, written for discount bonds with par value $100, as market convention suggests:

$$B(t_0,t_i) \quad \frac{100}{\Pi^{i=1}\,(1+\delta F(t_0,t_j)\,)} \qquad (23)$$

where the value of δ will be 1/2 in this example. Substituting the relevant forward rates from the preceding table, we indeed find that the given discount bond prices satisfy this equality. For example, for B (O, 2) ^{ask} we have

$$B(0,\,2)^{ask} \frac{100}{(1+.5(.0405))\;(1+.5(.0415))} \qquad (24)$$

We can obtain all forward rates, for the case δ = 1. By substituting the B (t_0, t_1) out from the first set of equations, we obtain n equations in n forward rates[12]. In the case of δ = ¼ or δ = ½, there are more unknown $F(t_0,t_i)$ than equations, if traded swap maturities are in years. Under these conditions the l_i would run over quarters whereas the superscript in s^n_{to},n = 1,2,............... will be in years. This is due to the fact that swap rates are quoted for annual intervals, whereas the settlement dates would be quaterly or semiannual. Some type of interpolation of swap rates or modeling will be required, which is common-even in traditional yield curve calculations.

3.2.Determining the $B(t_0,t_i)$ and the current Libor curve are provided by markets or

$$B(t_0,t_{i+1}) = \frac{1}{H^1_{3=0}\left(1+dF\left(t_0,t_j\right)\right)} \qquad (52)$$

are obtained from $[s^n_{to}]$ as in our case, we can use the formula to calculate the arbitrage-free values of the relevant pure dicount bond prices. In each case, we can derive the values of $B(t_0,t_{i+1})$ from the observed $\{F(t_0,t_i)\}$ and $\{s^n_{t0}\}$. This procedure would price the FRAs and bonds off the swap markets. It is called the curve algorithm.

3.3. Determining the Swap Rate

We can proceed in the opposite direction as well. Given arbitrage-free values of forward rates. we can, in principle, use the same formulas to determine the swap rates. All we need to do is (1) calculate the discount bond prices from the forward rates and (2) substitute these bond prices and the appropriate forward rates in our formula.

$$S^n_{t_0} = \frac{\sum_{i=0}^{n-1} B(t_0,t_{i+1})F(t_0,t_i)}{\sum_{i=0}^{n-1} B(t_0,t_{i+1})} \qquad (53)$$

Repeating this for all available S^n_{to}, n=1................30, we can obtain the arbitrage-free swap curve and discounts. In this case, we would be going from the spot and forward Libor curve to the (spot) swap curve.

3.4. Real-World Complications

There are, of course, several real-world complications to goint back and forth between the forward rates, discount bond prices, and swap rates. Let us mention three of these. First, as mentioned in the previous section, in reality swaps are taded for yearly intervals and the FRAs or Eurodollar contracts are traded for three-month or six-month or six-month tendors. This means that if we desired to go from swap quotes to quotes on forward rates using these formulas, there will be the need to interpolate the swap rates for portions of a year.

Second, observed quotes on forward rates do not necessarily come from paid-in-arrears FRAs, Market-traded FRAs settle at the time the Libor rate is observed, not at the end of the relevant period. The FRAs rates generated by these markets will be consistent with the formulas introduced earlier. On the other hand, some traders use interest rate futures, and, specifically, Eurocurrency futures, in hedging their swap books. Futures markets ar more transparent than the FRA market, and have a great deal of liquidity. But the forward rates determined in futures markets require convexity adjust ments before they can be used in the swap formulas discussed in this chapter.

3.4.1 Remark

Another important point needs to be mentioned here. Libor rates Li, apply to AA rated credits. This is implict in the fixing process of the BBA Libor. The banks that form the BBA panels have, in general, ratings of AA or AA-, and the interest rate that they pay reflects this level of credit risk. Our treatment has followed the general convention in academic work of using the term “Libor” as if it relates to a default-free loan.

Thus, if a financial engineer follows the procedures described here, the resulting curve will be the swap curve and not the treasury or sovereign curve. This swap curve will be “above” the sovereign or treasury curve, and the difference will be the curve for the swap spreads.

4. Term Structure Dynamics

In the remainder of this chapter, we will see that Forward Labor Model is the correct way to approach term structure dynamics. The model is based on the idea of converting the dynamics of each forward rate into a Martingale using some properly chosen forward measure. According to the linkages between sectors shown in this chapter, once such dynamics are obtained, we can use them to generate dynamics for other fixed-income instruments.

Most of the derivation associated with Forward Libor Model is an application of the fundamental theorem of asset pricing discussed in Chapter 11. Thus, we continue to use the same finite state world discussed in Chapter 11. The approach is mostly straightforward. There is only one aspect of forward Libor or swap models that makes them potentially difficult to follow. Depending on the instruments, arbitrage-free dynamics of different forward rates may have to be expressed under the same forward measure. The methodology then becomes more complicated. It requires a judicious sequence of Girsanov-style measure changes be applied to forward rate dynamics, in some recursive fashion. Otherwise, arbitrage-free dynamics of individual forward rates would not be correctly represented.

The Girsanov theorem is a powerful tool. But, it is not easy to conceive such successive measure changes. Doing this within a discrete framework, in a discrete setting, provides a great deal of motivation and facilities understanding of arbitrage-free dynamics. This is the purpose behind the second part of this chapter.

4.1. The Framework

We adopt a simple discrete framework and then extend it to general formulas. Consider a market where instruments can be priced and risk-managed in discrete times that are d apart.

$$t_o < t_1 < ..... < t_n = T \qquad (54)$$

with

$$t_i < t_{i-1} = d \qquad (55)$$

Initially, we concentrate on the first three times, t_o,t_1 and t_2 that are d apart. In this framework we consider four simple fixed-income securities:

- A default-free zero-coupon bond $B(t_o,t_1)$ that matures at time t_2 .
- A default-free zero-coupon bond that matres one period later, at time t_3. Its current price is expressed as $B(t_o,t_3)$.
- A savings account that pays (in-arrears) the discrete-time simple rate Lt, observed at time t_i. Therefore, the savings account payoff at t2 will be

$$R_{t_2} = (1+\delta L_{t_0})(1+\delta L_{t_1})$$

- Note that observed from the initial time to, the Lt1 will be a random variable.
- An FRA contracted at time to and settled at time t2, where the buyer receives/pays the differential between the fixed-rate $F(t_o,t_1)$ and the floating rate Lt1 at time t2. We let the notional amount of this instrument equal 1 and abbreviate the forward rate to Fto. The final payoff can be written as

$$(L_{t_1} - F_{t_0})\delta$$

These assets can be organized in the following payoff matrix D for time t2 as in Chapter 11, assuming that at every ti, from every node there are only two possible movements for the underlying random process. Denoting these movements by u,d, we can writen

$$D = \begin{bmatrix} R^{uu}_{t_2} & R^{ud}_{t_2} & R^{du}_{t_2} & R^{dd}_{t_2} \\ 1 & 1 & 1 & 1 \\ B^{uu}_{t_2} & B^{ud}_{t_2} & B^{du}_{t_2} & B^{dd}_{t_2} \\ \delta(F_{t_0} - L^u_{t_1}) & \delta(F_{t_0} - L^u_{t_1}) & \delta(F_{t_0} - L^d_{t_1}) & \delta(F_{t_0} - L^d_{t_1}) \end{bmatrix}$$

Where the Bij t2 is the random value of the t3 maturity discount bond at time t2. This value will be state-dependent at t2 because the bond matures one period later, at time t32. Looked at from time t0, this value will be random. Clearly, with this D matrix we have simplified the notation significantly. We are using only four states of the world, expressing the forward rate $F(t_o,t_2)$ simply as F t0, and the B(t2,t3) simply as B^{t3}_{t2}.

If the FRA, the savings account, and the two bonds do not admit any arbitrage opportunities, the fundamental theorem of asset pricing permits the following linear representation

$$\begin{bmatrix} 1 \\ B(t_0,t_2) \\ B(t_0,t_3) \\ 0 \end{bmatrix} = \begin{bmatrix} R^{uu}_{t_2} & R^{ud}_{t_2} & R^{du}_{t_2} & R^{dd}_{t_2} \\ 1 & 1 & 1 & 1 \\ B^{uu}_{t_2} & B^{ud}_{t_2} & B^{du}_{t_2} & B^{dd}_{t_2} \\ \delta(F_{t_0} - L^u_{t_1}) & \delta(F_{t_0} - L^u_{t_1}) & \delta(F_{t_0} - L^d_{t_1}) & \delta(F_{t_0} - L^d_{t_1}) \end{bmatrix} \begin{bmatrix} Q^{uu} \\ Q^{ud} \\ Q^{du} \\ Q^{dd} \end{bmatrix}$$

where $\{Q^{t3}$, i, j, u,d} are the four state prices for period t3. Under the arbitrage condition the latter exist and are positive.

$$Q^{t3} > 0 \qquad (60)$$

for all states i,j [14].

This matrix equation incorporates the ideas that (1) the fair market value of an FRA is zero at initiation, (2) 1 dollar is to be invested in the savings account originally, and (3) the bonds are default-free. They mature at times t_2 and t_3. The R^{t3} t_2, finally, represent the gross returns to the savings account as of time t_2. Because the interest rate that applies to time t1 is paid-in-arrears, at time t_{1+d}, we can express these gross returns as functions of the underlying Libor rates in the following way :

$$R^{uu}_{t_2} = R^{ud}_{t_2} = (1+\delta L_{t_0})(1+\delta L^u_{t_1})$$

$$R^{dd}_{t_2} = R^{du}_{t_2} = (1+\delta L_{t_0})(1+\delta L^d_{t_1})$$

We now present the Libor market model and the associated measure change methodology within this simple framework. The framework can be used to conveniently display most of the important tools and concepts that we need for fixed-income engineering. The first important concept that we need is the forward measure introduced in Chapter 11.

4.2. Normalization and Forward Measure is Inconvenient

To obtain the t2, and the t3 forward measures, it is best to begin with a risk-neutral probability, and show why it is not a good working measure in the fixed-income environment described earlier. We can then show how to convert the risk-neutral probability to a desired forward measure explicitly.

4.2.1 Risk-Neutral Measure is Inconvenient

As usual, define the risk-neutral measure {pij} using the first row of the matrix equation :

$$1 = R_{t_2}^{uu}Q^{uu} + R_{t_2}^{ud}Q^{ud} + R_{t_2}^{du}Q^{du} + R_{t_2}^{dd}Q^{dd}$$

$$\begin{aligned}\tilde{p}_{uu} &= R_{t_2}^{uu}Q^{uu}\\ \tilde{p}_{ud} &= R_{t_2}^{ud}Q^{ud}\\ \tilde{p}_{du} &= R_{t_2}^{du}Q^{du}\\ \tilde{p}_{dd} &= R_{t_2}^{dd}Q^{dd}\end{aligned}$$

The {Pij} then have the characteristics of a probability distribution, and they can be exploited with the associated Martingale equality.

We know from under the condition that every asset's price is argbitrage-free, {Qt3, i, j = u, d} exist and are all positive, and Pij will be the risk-neutral probabilities. Then, by using the last row of the system in Equation (59) we can write the following equality :

[redacted]

Here, (Ft0-Lt1), i = u,d are "normalized" so that Qij can be replaced by the respective pij. Note that in this equation, Fto is determined at time to, and can be factored out. Grouping and rearranging, we get

[redacted]

This can be written using the expectation operator

$$F_{t_0} = \frac{1}{E_{t_0}^{\tilde{P}}\left[\frac{1}{R_{t_2}}\right]} E_{t_0}^{\tilde{P}}\left[\frac{L_{t_1}}{R_{t_2}}\right]$$

According to this last equality, if Rt2 is a random variable and is not independent of Lt1, 15 it cannot be moved outside the expectation operator. In other words, for general t,

$$F(t, t_i) \neq E_t^{\tilde{P}}[L_{t_i}] \quad t < t_i$$

That is to say, under the risk-neutral measure, P, the forward rate for time ti is a biased "forecast" of the future Libor rate Lt. In fact, it is not very difficult to see that the futures rate that will be determined by, san, a Curodollar contract at time t. The "bias" in the forward rate, therefore, is associated with the convexity adjustment.

Another way of putting it is that, F_t is not a Martingale with respect to the risk –neutral probability P, and that a discritized stochastic difference equation that represents the dynamics of F_t will, in general, have a trend :

[redacted]

where a(Ft,t) is the nonzero expected rate of change of the forward rate under the probability P.

The fact that Ft, is not a Martingale with respect to probability P makes the risk-neutral measure an inconvenient working tool for pricing and risk management in the fixed-income sector. Before we can use Equation 72 we need to calibrate the drift factor a(.). This requires first obtaining a functional form for the drift under the probability P. The original HJM article does develop a functional form for such drifts using continously compounded instantaneous forward rates. But, this creates an environment quite different from Libor-driven markets and the associated acturial rates Lt, used here.

On the other hand, we will see that in the interest rate sector, arbitrage-free drifts become much easier to calculate if we use the Forward Libor Model and switch to appropriate forward measures.

4.2.2 The Forward Measure

Consider defining a new set of probabilities for the sates under consideration by using the default-free zero-coupon bond that matures at time t2. First, we present the simple case. Use the second row of the system in Equation (59).

$$B(t_0, t_2) = Q^{uu} + Q^{ud} + Q^{du} + Q^{dd}$$

and then divide every element by $B(t_0, t_2)$. Renaming, we get the forward t_2-measure Pt_2.

$$1 = \tilde{p}_{uu}^{t_2} + \tilde{p}_{ud}^{t_2} + \tilde{p}_{du}^{t_2} + \tilde{p}_{dd}^{t_2}$$

where the probability of each state is obtained by scaling the corresponding Q^{ij} using the time t0 price of the corresponding bond :

$$\tilde{p}_{ij}^{t_2} = \frac{Q^{ij}}{B(t_0, t_2)}$$

It is important to index the forward measure with the superscript, t2, in these fixed-income models, as other forward measures would be needed for other forward rates. The superscript is a nice way of keeping track of the measure being used. For some instruments, these measures have to be switched sequentially.

Using the t_2 -forward measure we can price any asset C_t, with time-t_2, payoffs C^{ij}_{t2}.

$$C_{t_0} = [B(t_0, t_2)C_{t_2}^{uu}]\tilde{p}_{uu}^{t_2} + [B(t_0, t_2)C_{t_2}^{ud}]\tilde{p}_{ud}^{t_2}$$

$$+[B(t_0,t_2)C^{du}_{t_2}]\tilde{p}^{t_2}_{du}+[B(t_0,t_2)C^{dd}_{t_2}]\tilde{p}^{t_2}_{dd}$$

This implies that, for an asset that settles at time T and has no other payouts, the general pricing equation is given by

$$C_t = B(t,T)E^{\tilde{P}^T}_t[C_T]$$

where P^T is the associated T-forward measure and where C_T is the time-T payoff. According to this equality, it is the ratio

$$Z_t = \frac{C_t}{B(t,T)}$$

which is a Martingale with respect to the measure P^T. In fact, B(t,T) being the discount factor for time T, and, hence, being less than one, Z_t is nothing more than T-forward value of the C_t. This means that the forward measure PT operates in terms of Martingales that are measured in time-T dollars. The advantage of the forward PT measure becomes clear if we apply the same transformation to price the FRA as was done earlier for the case the of risk-neutral measure.

4.2.3 Arbitrage-Free SDEs for Forward Rates

To get arbitrage-free dynamics for forward rates, we now go back to the simple model in Equation (59). Dividing the fourth now of the system by $B(t_o, t_2)$ and rearranging.

$$\frac{F_{t_0}}{B(t_0,t_2)}[Q^{uu}+Q^{ud}+Q^{du}+Q^{dd}] =$$

$$\frac{L^u_{t_1}}{B(t_0,t_2)}Q^{uu}+\frac{L^u_{t_1}}{B(t_0,t_2)}Q^{ud}$$
$$+\frac{L^d_{t_1}}{B(t_0,t_2)}Q^{du}+\frac{L^d_{t_1}}{B(t_0,t_2)}Q^{dd}$$

substitute the t_2-forward measure, into this equation using the equality :

$$\tilde{p}^{t_2}_{ij} = \frac{1}{B(t_0,t_2)}Q^{ij}$$

The equation becomes

$$F_{t_0} = [L^u_{t_1}]\tilde{p}^{t_2}_{uu}+[L^u_{t_1}]\tilde{p}^{t_2}_{ud}+[L^d_{t_1}]\tilde{p}^{t_2}_{du}+[L^d_{t_1}]\tilde{p}^{t_2}_{dd}$$

Extending this to the general case of m discrete states

$$F_{t_0} = \sum_{i=1}^{m} L^i_{t_1}\tilde{p}^{t_2}_i$$

This is clearly the expectation

$$F_{t_0} = E^{\tilde{P}^{t_2}}_{t_0}[L_{t_1}]$$

This means that, under the measure Pt_2, the forward rate for the period $[t_1,t_2]$ will be unbiased estimate of the corresponding Libor rate.

Consequently, switching to the general notation of (t,T), the process

$$F_t = T(t,T,T+\delta) \quad (83)$$

will be a Martingale under the (T + δ) – forward measure P^{T+d}. Assuming that the errors due to discretization are small, its dynamics can be described by a (discretized) SDE over small intervals of length Δ[17].

$$F_{t+\Delta} - F_t = \sigma_t F_t \Delta W_t \quad (84)$$

where Wt is Wiener process under the measure P^{T+d}. DW_t is the Wiener process increment :

$$\Delta W_t = W_{t+\Delta} - W_t \quad (85)$$

This (approximate) equation hs a no drift component since, by arguments, and writing for the general t. T, we have

$$1 + \delta F(t,T) = \frac{B(t,T)}{B(t,T+\delta)} \quad (86)$$

It is clear from the normalizaton arguments of Chapter 11 that, under the measure P^{T+d} and normalization by B(t,T + δ) , the ration on the right-hand side of this equation is a Martingale with respect to P^{t+d}. this makes the corresponding forward rate a Martingale, so that the implied SDE will have no drift.

However, note that the forward rate for the period (T –δ, T) given by

$$1+\delta\ F(t,T-\delta) = \frac{B(t,T-\delta)}{B(t,T)} \quad (87)$$

is not a Martingale under the same forward measure $P^{T+\delta}$ Instead, this forward rate is a Martingale under it own measure P^T which requries normalization by B (t,T). Thus, we get a critical result for the Forward Libor Model :

Each forward rate F(t,T), admits a Martingale representation under its own forward measure $P^{Ti+\delta}$

This means that each forward rate dynamics can be approximated individually by a difference equation with no drift given the proper normalization. The only parmaeter that would be needed to characterize such dynamics is the coressponding forward rate volatility.

Arbitrage-Free Dynamics

The previous section discussed the dynamics of forward rates under their own forward measure. We now show what happens when we use one forward measure for two forward rates that apply to two consecutive periods. Then, one of the forward rates has to be evaluated under a measure different from its own, the Martingale dynamics will be broken. Yet, we will be able to obtain the new drift.

To keep the issue as simple as possible, we continue with the basic model in Equation (59), except, we add one more time period so that we can work with two non-trivial forward rates and their respective forward measures. This is the simplest setup within which we can show how measure-change technology can be implemented. Using the forward measures introduced earlier and shown in Fig. 13-4, we can now define the following forward rate dynacmics for the tow forward Libor processes {F (t0, t1), F(t0, t2)} under consideration. The first will be a Martingale under the normalization with B (t0, t2). In our simplified setup, we will observe only two future values of this forward rate at times t1 and t2. These are given by

$$F(t_1, t_2) - F(t_0, t_2) = \sigma_2 F(t_0, t_2) \Delta W^{t_3}_{t_1}$$

$$F(t_2, t_2) - F(t_1, t_2) = \sigma_2 F(t_1, t_2) \Delta W^{t_3}_{t_1}$$

The superscript in Wt3, i = 1, 2 indicates that the Wiener process increments have zero mean under the probability Pt_3. These equations show how the "current" value of the forward rate F (t_0, t_2) first changes to become $F(t_1, t_2)$ and then ends up as $F(t_2, t_2)$. The latter is also Lt_2.

For the "near' forward rate $F(t_0, t_1)$, we need only on equation[18] defined under the normalization with the bond B (t_0, t_1) (i.e, th Pt_2 measure) and the associated zero drift.

$$F(t_1, t_1) - F(t_0\, t_1) = \sigma_1 F(t_0, t_1) \Delta W^{t_2}_{t1}$$

Similarly, the superscript in Wt_2t_1 indicates that this Wiener process increment has zero mean under the probability Pt_2. Here, the $F(t_1, t_2)$ is also the Libor rate Ltv. We reemphasize that each dynamics is defined under a different probability measure. Under these different forward measures. each forward Libor process behaves like a Martinglae.[19]. Consequently, there are no drift terms in either equation.

Fortunately, as long as we can work with these equations separately, no arbitrage-free drift terms need to the estimated or calibrated. The only parameters we need to determine are the volatilities of the two forward rates ; s_2 for the forward rate $F(t_0, t_2)$, and σ_1 for the forward rate F(t., t1)[20].

In fact, each Wiener increment has a zero expectation under its own measure. For example, the Wiener increments of the two forward rates will satisfy, for time $t_0<t_1$.

$$E^{Pt2}_{t0} [\Delta W^{t2}_{t1}] = 0 \qquad 90$$

$$E^{Pt3}_{t0} [\Delta W^{t3}_{t1}] = 0 \qquad 91$$

[18] This forward rate process will terminate at t_2.

[19] Again, we are assuming that the discretization bias is negligible.

[20] Note that according to the characterization here, the volatility parameters are not allowed to vary over time. This assumption can be relaxed somewhat but we prefer the simple setting. Since most market applications are based on constant volatility, characterzation as well

Yet when we evaluate the expectations under Pt^2, we get

$$E^{Pt2}_{t0} [\Delta W^{t2}_{t1}] = 0 \qquad 92$$

$$E^{Pt2}_{t0} [\Delta W^{t2}_{t1}] = 0 \qquad 93$$

Here is a mean correction that needs to be made because we are evaluating the Wiener increment under a measure different from its own forward measure P^{t3}. This, in turn, means that the dynamics for $F(t_0, t_1)$ lose its Martingale characterstic.

We will now comment on the second moments, variances and covariances. Each Weiner increment is assumed to have the same variance under the two measures. The Girsanov theorm ensures that this is true in continuous time. In discrete time, this holds as an approximation. Finally, we are operating in an environment where there is only factor.21. So, the Wiener process increments defined under the two forward measures will be exactly correlated if they belong to the same time period. In other word, although their means are different. We can assume that, approximately, their covariance would be Δ

$$E^{Pt3} [\Delta W^{t3}_{t} \quad \Delta W^{t2}_{t}] = EP^{t2} [\Delta W^{t3}_{t} \ \Delta Wt^{2}_{t}] = \Delta$$

Similar equalities with hold for the variances as well 22.

Review

The results thus far indicate that for the pricing and risk managing of equity-linked assets. the risk-neutral measure P may be quite convient since easily adaptable to lognormal models where the arbitrage free drifts are simple and known functions of the risk-free interest rate. As far as equity products are concerned, the assumption that short rates are constant is a reasonable approximation, especially for short maturities. yet, for contracts written on future values of interest rates (rather than on asset prices), the use of the P leads to complex arbitrage free dynamics that cannot be captured easily by Martingales and, hence the corresponding arbitrage-free drift terms may be difficult to calibrate.

Approximate forward measures, on the other hand, result in martingale equalities and lead to dynamics convenient for the calculation of arbitrage-free-drifts, even when they are not zero. Forward (and swap) measures are the proper working probabilities for fixed- income environments.

A Monte Carlo Implementation

Suppose we want to generate Monte Carlo "paths" from the two discretized SDEs for two forward rates, $F(t_1, t_2)$.

$$F(t_i, t_1) - F(t_i - 1, t_1) = \sigma_1 F(t_i - 1, t_1) \Delta W t_{t_i}^2 \qquad 94$$

$$F(t_i, t_2) - F(t_{i-1}, t_2) = \sigma_2 F(t_{i-1}, t_2) \Delta W t_{t_i}^3 \qquad 95$$

where i =1,2 for the second equation, and i=1 for the first.

[21] As a reminder, a one-factor model assumes that all random processes under consideration have the same unpredictable component up to a factor proportionality. In other words, the correlation coefficients between these processes would be one.

[22] These relations will hold as Δ goes to zero.

It is easy to generate individual paths for the two forward rates separetely, by using these Martingale equations defined under their own forward measures. Consider the following approach. Suppose volatilities 1 and 2 can be observed in the market. We select two random variables {DW3, DW3 } from the distribution

$$DW_i^3 \sim N(0, \Delta) \qquad 96$$

with a pseudo-random number generator, and then calculate, sequentially, the randomly-generated forward rates in the following order, starting with the observed F(t0, t2)

$$F(t_1\text{-}t_2)^1 = F(t_0, t_2) + \sigma_2 F(t_0, t_2) \ \Delta W_1^3$$

$$F(t_2\text{-}t_2)^1 = F(t_1, t_2) + \sigma_2 F(t_1, t_2) \ \Delta W_2^3$$

where the superscript on the left-hand side indicates that these values are for the first Monte Carlo trajectory. Proceeding sequentially, all the terms on the right-hand side will be known. This gives the first simulated "path" {F(t0, t2), F(t1, t2)1, F(t2, t2)1}. we can repeat this algorithm to obtain M such paths for potential use in pricing.

What does this imply for the other forward Libor process F(t., t1) ? Can we use the same randomly, generated random variable W3 in the Martingale equation for F (t, t1) and obtain the first "path" {F (t0, t1), F(t1, t1) 1 } from

$$F(t_1\text{-}t_1)^1 = F(t_0, t_1) + \sigma_1 F(t_0, t_1) \quad \Delta W_1^3$$

The answer is no. As mentioned earlier, the Wiener increments {Wt2} have zero mean only under the probability Pt2, but, the first set of random variables was selected using the measure Pt3, Under Pt2, these random varibles do not have zero mean, but are disturbted as

$$N(\lambda_{t_0}^{t_2}\Delta, \Delta)$$

Thus, if we use the same W3 in Equation (99), then we need to correct for the term . To do this, we need to calculate the numerical value

$$F(t_1, t_1) = F(t_0, t_1) - \sigma_1 F(t_0, t_1)(\lambda_{t_0}^{t_2}\Delta) + \dot{\sigma}_1 F(t_0, t_1)\Delta W_{t_1}^{t_3}$$

To see why this is to, take the expectation under Pt_2 on the right-hand side and use the information in Eq (100):

$$E_{t_1}^{\tilde{P}^{t_2}}[F(t_0, t_1) - \sigma_1 F(t_0, t_1)(\lambda_{t_0}^{t_2}\Delta) + \sigma_1 F(t_0, t_1)\Delta W_{t_1}^{t_3}] = F(t_0, t_1)$$

Thus, we get the correct result under the Pt2, after reh mean correction. It is obvious that we need to determine these correction factors before the randomly generated Brownian motion increments can be used in all equations.

Yet, notice the following simple case. If the instrument under consideration has additive cash flows where each cash flow depends on a single forward rate, then individual zero-drift equations can be used separately to generate paths. This applies fro several liquid instruments. for example FRAs and especially swaps have payment legs that depend on one Libor rate only. Individual zero-drift equations can be used for valuting each leg separately, and then these values can be added using observed zero coupon bond prices B(t, T,). However, this cannot be done in the case of constant maturity swaps for example, because each settlement leg will depend nonlinearly on more than one forward rate.

We now discuss further how mean corrections can be conducted so that all forward rates are projected using a single forward measure. This will permit pricing instruments where individual cash flows depend on more than one forward rate.

5. Measure Change Technology

We introduce a relatively general framework and then apply the results to the simple example shown previously. Basically, we need three previously-developed relationships. We let t_i obey

$$t_o < \ldots < t_n = T$$

with

$$t_i\text{-}t_{i-1} = \delta$$

denote settlement dates of basic interest rate swap structure and limit our attention to forward rates for successive forward loans contracted to begin at t_{i+1}. An example is shown in Figure 13-4.

Result 1

The forward rate at time t, for a Libor-based forward loan that starts at time t_i and ends at time ti+d, denoted by F (t, t_i), admits the following arbitrage relationship :

$$1+F(t,t_i)\, d = \frac{B(t, t_i)}{B(t,t_{i+1})} \qquad t < t_i \qquad (102)$$

where, as usual, B(t, t_i) and B(t, t_{i+1}) are the time-t prices of default –free zero coupon bonds that mature at times t_i and t_{i+1}, respectively.

The left side of this equality is a gross forward return. The right side, on the other hand, is a treated asset price, B(t, t_i), normalized by another asset price, the B(t, t_{i+1}). Hence, the ratio will be a Martingale under a proper measure, here, the forward measure denoted by P^{i+1}.

Result 2

In a discrete state setting wit k states of the world, assuming that all asset prices are arbitrage-free, and that time-t_i state prices Q^j, j = 1, . . . k, with 0 < Q^j exist, the time-t_i values to the forward $P^{\sim ti}$ are given by[23]

$p_1^{\sim ti} =$	$\frac{1}{B(t,t_i)}$	$Q^1, p_2^{\sim ti} =$	$\frac{1}{B(t,ti)}$	$Q^2, \ldots P^{ti} =$	$\frac{1}{B(t,ti)}$	Q^k (103)

These probabilities satisfy :

$p_1^{\sim ti} + p_2^{\sim ti} + \ldots p_k^{\sim ti} = 1$ (104)

and

$0 < p_1^{\sim ti}$ "j

Note that the proportionality factors used to convert Q^j into $p_j^{\sim ti}$ are equal across j.

Result 3

In the same setting, the time t_i probabilities associated with the t_{i+1} forward measure $p_1^{\sim ti+1}$ are given by :

$p_1^{\sim ti+1} =$	$\frac{B(t_i,t_{i+1})^1}{B(t,t_{i+1})}$	$Q^1, p_2^{\sim ti+1} =$	$\frac{B(t_i,t_{i+1})^2}{B(t,t_{i+1})}$	$Q^2, \ldots, pk^{\sim ti+1} =$	$\frac{B(t_i,t_{i+1})^k}{B(t,t_{i+1})}$	Q^k (105)

where the $B(t_i,t_{i+1})^j$ are the state dependent values of the t_{i+1} maturity bond at time t_i. Here, the bond that matures at time ti+1 is used to normalize the cash flows for time t_i. Since the maturity date is t_{i+1}, the $B(t_i, t_{i+1})^j$ are not constant at t_i. The factors used to convert $\{Q^j\}$ into $p_1^{\sim ti+1}$ cease to be constant as well.

We use these results in discussing the mechanics of measure changes. Suppose we need to price an instrument whose value depends on two forward Libor processes, $F(t, t_{i+1})$ and $F(t, t_{i+1})$, simultaneously. We know that each process is a Martingale and obeys an SDE with zero-drift under its own forward mearure.

Consider a one-factor setting, where a single Wiener process cause fluctuations in the two forward rates. Suppose that in this setting, starting from time t, with $t < t_i$, i=1, . . . n, a small time interval denoted by h passes with $t+h < t_i$. By imposing a Gaussian volatility structure, we can write down the individual discretized arbitrage-free dynamics for two successive forward rates $F(t, t_i)$ and $F(t, t_{i+1})$ as

$$F(t+h, t_i) - F(t, t_i) = \sigma^i F(t, t_i)\Delta W^1_{t+h} \quad (106)$$

and

$$F(t+h, t_{i+1}) - F(t, t_{i+1}) = \sigma^{i+1} F(t, t_{i+1})\Delta W^2_{t+h} \quad (107)$$

Changes in these forward rates have zero mean under their own forward measure and, hence are written wit zero drift. This means that the unique real world Wiener process W_{t+h} is now denoted by ΔW^1_{t+h} and ΔW^2_{t+h} in the two equations. These are normally distributed, with mean zero and variance h only under their own forward measures, t_{i+1} and t_{i+2} forward probability measures, respectively.24 Finally, note how we simplify the characterization of volatilities and assume that they are constant over time.

The individual Martingale dynamics are very convenient from a financial engineering point of view. The respective drift components are zero and, hence, they need not be modeled during pricing. The only major task of the market practitioner is to get the respective volatilities σ^1 and σ^{i+1}.

However, some securities prices may depend on more than one forward rate in a non-linear fashion and their value may have to be calculated as an expectation under one single measure. For example, suppose a security's price, St, depends on F(t,ti) and F(t, t i+1)through a pricing relation such as :

where g(.) is a known non-linear function. Then, the expectation has to be calculated under one measure only. This probability can be either the time-t_{i+1}, or the time-t_{i+2} forward measure.

We then have to choose a forward rate equation with martingale dynamics and carry out a mean correction to get the correct arbitrage-free dynamics for the other. The forward measure of one of the Martingale relationships is set as the working probability distribution, and the other equation(s) is obtained in terms of this unique probability by going through successive measure changes. We discuss this in detail below.

5.1 The Mechanics of Measure Changes

We have the following expectations concerning ΔW^1_{t+h} and DW^2_{t+h}, defined in (106) and (107)

$$E_t^{\tilde{P}^{t_{i+1}}}\left[\Delta W^1_{t+h}\right] = 0$$
$$E_t^{\tilde{P}^{t_{i+2}}}\left[\Delta W^2_{t+h}\right] = 0$$

Under their own forward measure, each Wiener increment has zero expectation. If we select $P^{t\,i+2}$ as our working measure, one of these equalities has to change. We would have

$$E_t^{\tilde{P}^{t_{i+2}}}\left[\Delta W^1_{t+h}\right] = \lambda_t h$$
$$E_t^{\tilde{P}^{t_{i+2}}}\left[\Delta W^2_{t+h}\right] = 0$$

The value of lt gives the correction factor that we need to use in order to obtain the correct arbitrage-free dynamics, if the working measure is $P^{t\,i+2}$. Calculating this factor implies that we can change measures in the dynamics of $F(t, t_i)$.

We start with the original expectation :

$$E_t^{\tilde{P}^{t_{i+1}}}\left[\Delta W^1_{t+h}\right] = \sum_{j=1}^{k} \Delta W^{1j}_{t+h} \tilde{p}_j^{t_{i+1}} = 0$$

where the $P_j^{t\,i+1}$ are the probabilities associated with the individual states j=1.........k. Now, using the identity.

$$\frac{B(t,t_{i+2})B(t_i,t_{i+2})^j}{B(t,t_{i+2})B(t_i,t_{i+2})^j} \equiv 1$$

We rewrite the expectation as

$$E_t^{\tilde{P}^{t_{i+1}}}\left[\Delta W^1_{t+h}\right] = \sum_{j=1}^{k}(\Delta W^{1j}_{t+h})\left[\frac{B(t,t_{i+2})B(t_i,t_{i+2})^j}{B(t,t_{i+2})B(t_i,t_{i+2})^j}\right]\tilde{p}_j^{t_{i+1}}$$

We regroup and use the definition of the t i+1 and t i+2 forward measures and implied by Result 3

$$\tilde{p}_j^{t_{i+1}} = \frac{B(t_i,t_{i+1})^j}{B(t,t_{i+1})}Q^j$$

and

$$\tilde{p}_j^{t_{i+2}} = \frac{B(t_i, t_{i+2})^j}{B(t, t_{i+2})} Q^j$$

Re-scaling the Q3 using appropriate factors, (115) becomes

$$\sum_{j=1}^{k} (\Delta W_{t+h}^{1j}) \left[\frac{B(t, t_{i+2})}{B(t, t_{i+1})} \frac{B(t_i, t_{i+1})^j}{B(t_i, t_{i+2})^j} \right] \tilde{p}_j^{t_{i+2}} = 0$$

Note that the probabilities switch as the factors that were applied to the Q^3 changed. The superscript in W^1_{t+h} does not change.

The next step in the derivation is to try to "recognize" the elements in this expectation. Using Result 1, we recognise the equality.

$$1 + \delta F(t_i, t_{i+1})^j = \frac{B(t_i, t_{i+1})^j}{B(t_i, t_{i+2})^j}$$

Replacing, eliminating the j-independent terms, and rearranging gives

$$\sum_{j=1}^{k} (\Delta W_{t+h}^{1j})(1 + \delta F(t_i, t_{i+1})^j) \tilde{p}_j^{t_{i+2}} = 0$$

Now, multiplying through, this leads to

$$\sum_{j=1}^{k} (\Delta W_{t+h}^{1j}) \tilde{p}_j^{t_{i+2}} = -\left(\sum_{j=1}^{k} (\Delta W_{t+h}^{1j}) F(t_i, t_{i+1})^j \tilde{p}_j^{t_{i+2}} \right) \delta$$

We can write this using the conditional expectation operator.

$$E_t^{\tilde{P}^{t_{i+2}}} \left[\Delta W_{t+h}^1 \right] = -E_t^{\tilde{P}^{t_{i+2}}} \left[\Delta W_{t+h}^1 F(t_i, t_{i+1}) \right] \delta.$$

In the last expression, the left-hand side is the desired expectation of the ΔW^1_{t+h} under the new probability $P^{t_{i+2}}$. This expectation will not equal zero if the right-hand side random variables are correlated. This correlation is non-zero as long as forward rates are correlated. To evaluate the mean of ΔW^1_{t+h} under the new probability $P^{t_{i+2}}$, we then have to calculate the covariance.

Let the covariance be given by $-l_t$ h. We have,

$$\delta E_t^{\tilde{P}^{t_{i+2}}} \left[\Delta W_{t+h}^1 F(t_i, t_{i+1}) \right] = -\lambda_t h$$

Using the lt we can switch probabilities in the F(t,ti) dynamics. We start with the original Martingale dynamics.

$$F(t+h, t_i) = F(t, t_i) + \sigma^i F(t, t_i) \Delta W_{t+h}^1$$

Switch by adding and subtracting si $F(t,ti)l_t$ h to the right-hand side and regroup :

$$F(t+h, t_i) = F(t, t_i) - \sigma^i F(t, t_i) \lambda_t h$$

$$+ \Delta W_{t+h}^1]$$

Let

$$\Delta W_{t+h}^2 = [\lambda_t h + \Delta W_{t+h}^1]$$

We have just shown that the expectation of the right-hand side of this expression equals zero under Pt^{i+2}. So, under the Pt^{i+2} we can write the new dynamics of the F(t,t) as

$$F(t+h, t_i) = F(t, t_i) - \sigma^i F(t, t_i) \lambda_t h$$

$$+ \sigma^i F(t, t_i) \Delta W_{t+h}^2$$

As can be seen from this expression, the new dynamics have a non-zero drift and the F(t,t) is not a Martingale under the new measure. Yet, this process is arbitrage-free and easy to exploit in Monte-Carlo type approaches. Since both dynamics are expressed under the same measure, the set of equations that describe the dynamics of the two forward rates can be used in pricing instruments that depend on these forward rates. The same pseudo-random numbers can be used in the two SDEs. Finally, the reader should remember that the discussion in this section depends on the discrete approximation of the SDEs.

5.2. Generalization

A generalization of the previous heuristic discussion leads to the Forward Libor Model. Suppose the setting involves n forward rates, F(to,ti), i = 0,..............n-1, that apply to loans which begin at time ti, and end at ti+1 = ti+d. The F(to,to) is the trivial forward rate and is the spot Libor with tenor d. The terminal date is tn.

Similar to the discussion in the previous section, assume that there is a single factor26. Using the ti+1 forward measure we obtain arbitrate-free Martingale dynamics for each forward rate F(t,ti) :

$$dF(t, t_i) = \sigma^i F(t, t_i) W_t^{i+1} \quad t \in [0, \infty)$$

The superscript in Wt i+1, implies that 27

$$E_t^{\tilde{P}^{t_{i+1}}} [dW_t^{i+1}] = 0$$

These arbitage-free dynamics are very useful since they do not involve any interest rate moeling and are dependent only on the correct specification of the respective volatilities. However, when more than one forward rate determines a security's payoff in a non-linear fashion, the process may have to be written under a unique working measure.

Suppose we chose Ptn as the working measure. The heuristic approach discussed in the previous section can be generalized to obtain the following arbitrage –free system of SDEs that involve recursive drift corrections in one-factor case :

$$dF(t,t_i) = -\left[\sigma^i F(t,t_i)\sum_{j=i}^{n-1}\frac{\delta\sigma^j F(t,t_j)}{1+F(t,t_j)\delta}\right]$$

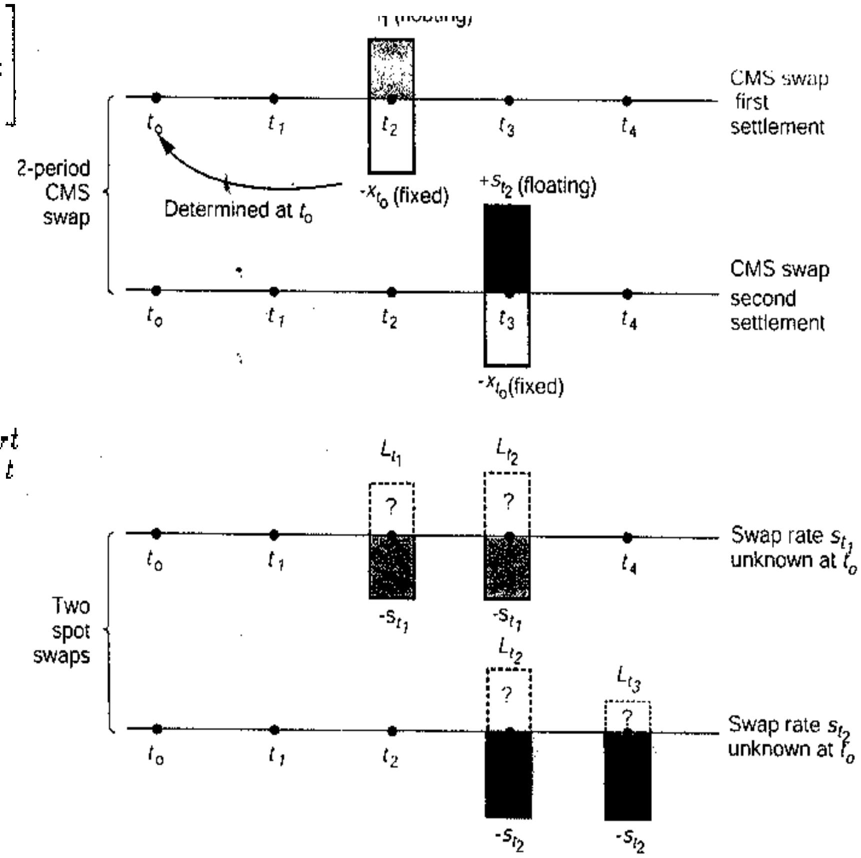
$$dt + \sigma^i F(t,t_i)dW_t^{t_n} \qquad t \in [0,\infty)$$

where the superscript in the dWltn indicates that the working measure is Ptn. The equations in this system are expressed under this forward measure for i =1 .... n. Yet, only the last equation has a Martingale dynamics.

$$dF(t,t_{n-1}) = \sigma^{n-1}F(t,t_{n-1})dW_t^t$$

$$t \in [0,\infty)$$

All other SDEs involve successive correction factors given by the first term on the right side. It is important to realize that all terms in these factors can be observed at time t. The dynamics does not need a modeling of actual drifts.



6. An Application

The forward measure change technology are relevant for the pricing of many instruments. But there is one instrument class that has recently become quite popular with market participants and that can be priced with this technology. These are constant maturity swaps (CMS). They have properties that would illustrate some subteties of the methods used thus far. In order to price them, forward rates need to be projected jointly.

First, we present a reading that illustrated some of the recent interest in this instrument class.

Example

Institutional investors, convinced that euro-zone interest rates are about to rise, have over the part month hoovered up over US $ 4 bn of notes paying coupons linked to constant maturity swap (CMS) rates. Swelling demand for these products could resusicitate the ailing market in step-up callable bonds and lead to a longer –term balance in European options markets. The CMS boom is being driven by European institutional investors keen to speculate on higher European interest rates.

The CMS deal structure is fairly generic and similar engineering was in evidence earlier in 1999. The Italy issue is typical, offering investors a 4% coupon in year one and 78% of the 10-year CMS rate for the remaining 19 years. Most deals include a floor limiting the investor's downside coupon rate.

CMS-based products appear very attractive in the current yield curve environment. They offer an above market fist coupon and the chance to speculate on rising interest rates. They also guarantee a minimum coupon of at least 4% (IFR, Issue 1281)

CMS swaps are instruments that build on the plain vanilla swaps in an interesting way. In a vanilla swap, a fixed swap rate is exchanged against a floating Libor that is an interest rates relevant for that particular settlement period only. In a CMS swap, this will be generalized. The fixed leg is exchanged against a floating leg, but the floating leg is not a "one-period" rate. It is itself a multi-period swap rate that will be determined in the future.

There are many versions of such exchanges, but as an example we consider the followiing. Suppose one party decides to pay 4% during the next three years against receiving a 2-year swap rate that will be determined at the beginning of each one of those years. The future swap rates are unknown at time to, and can be considered as floating payments. Except, they are not floating payments that depend on the perceived volatility for that particular year only. They are themselves averages of one-year rates. Clearly, such swaps have significant nonlinearities and we cannot do the same engineering as in the case of a plain vanilla swap.

An example of CMS swaps is shown in Figure 13-5. The reader can see that what is being exchanged at each settlement date against a fixed payment is a floating rate that is a function of more than one forward rate. Under these conditions it is impossible to project individual forward rates using individual zero-drift stochastic differential equations defined under different forward measures. Each leg of the CMS swap depends on more than one forward rate and these need to be projected jointly, under a single measure.

1.6. Another Example of Measure Change

This section provides another example to measure change technology from the FRA markets. Paid-in-appears FRAs make time-t_{i+1} payoffs :

$$N\delta[F(t_0, t_i) - L_{t_i}]$$

The market-traded FRAs, on the other hand, settle at time ti according to :

$$\frac{N\delta[F(t_0, t_i) - L_{t_i}]}{(1 + \delta L_{t_i})}$$

Finally, we have Libor-in-arrears FRAs that settle according to

$$N\delta[F(t_0, t_i) - L_{t_i}]$$

at time ti. As we have seen in Chapter 9, the Libor-in-arrear FRA payoffs settle in a "non-natural" way, since L_{t_i}- related payments would normally be received or paid at time t_{i+1}.

we now show that the paid-in-arrears FRA and market-traded FRAs lead to the same forward rate. First, remember that under the $P^{t_{i+1}}$ forward measure for paid-in-arrears FRAs, we have :

$$F(t_0, t_i) = E_{t_0}^{\tilde{P}^{t_{i+1}}}[L_{t_i}]$$

That is to say, the FRA rate F(to,ti) is the average of possible values the Libor rate might take :

$$F(t_0, t_i) = \sum_{j=1}^{k} L_{t_i}^{j} \tilde{p}_j^{t_{i+1}}$$

where j represents possible states of the world, which are assumed to be discrete and countable.

Now, consider the settelement amount of market-traded FRAs:

$$\frac{N\delta[F(t_0, t_i) - L_{t_i}]}{(1 + \delta L_{t_i})}$$

Would the forward rate implied by this contract be the same as the paid-in-arrears FRAs ?

The answer is yes. Using the measure change Technology, we discuss how this can be shown. The idea is to begin with the expectation of this settelement amount under teh Pti measure, and show that it leads to the same forward rate. Thus, begin with

$$E_{t_0}^{\tilde{P}^{t_i}}\left[\frac{N\delta[F(t_0, t_i) - L_{t_i}]}{(1 + \delta L_{t_i})}\right]$$

Setting this equal to zero, and rearranging, leads to the pricing equation

$$F(t_0, t_i) = \frac{E_{t_0}^{\tilde{P}^{t_i}}\left[\frac{N\delta L_{t_i}}{(1+\delta L_{t_i})}\right]}{E_{t_0}^{\tilde{P}_{t_i}}\left[\frac{N\delta}{(1+\delta L_{t_i})}\right]}$$

Now we switch on the right-hand side of Equation (130). We have two expectations and we shall switch measures in both of them. But first, let N =1 and similarly d =1.

Consider the numerator

$$E_{t_0}^{\tilde{P}^{t_i}}\left[\frac{L_{t_i}}{(1 + L_{t_i})}\right] = \sum_{j=1}^{k} \frac{L_{t_i}^{j}}{(1 + L_{t_i}^{j})} \tilde{p}_j^{t_i}$$

We know that for time t_i

$$\tilde{p}_j^{t_i} = \frac{1}{B(t_0, t_i)} Q^j$$

$$\tilde{p}_j^{t_{i+1}} = \frac{B(t_i, t_{i+1})^j}{B(t_0, t_{i+1})} Q^j$$

Thus :

$$\tilde{p}_j^{t_i} = \frac{1}{B(t_0, t_i)} \frac{B(t_0; t_{i+1})}{B(t_i, t_{i+1})^j} \tilde{p}_j^{t_{i+1}}$$

Note that, again the random $(1+L^j_{ti})$ terms conveniently cancel, and on the right-hand side we obtain :

Putting the numerator and denominator together for general N and δ gives

We simplify the common terms to get

$F(t_0, t_1) - E^{Pti+1}$ [Lj ti]

Hence, we obtained the desired result. The FRA rate of paid-in-arrears. FRAs is identical to the FRA rate of market-traded FRAs and is an unbiased predicator of the Libor rate Lti, under the right forward measure.

We conclude this section with another simple example.

Example

We can apply the forward measure technology to mark-to-market practices as well. The paid-in-arrears FRA will settle at time ti+1 according to

$$[L_{ti} - F(t_0, t_i)]\ N$$

What is the value of this contract at time t1, with $t_o < t_1 < t_i$?

It is market convention to replace the random variable Lt, with the corresponding forward rate of time t_1. We get

$$[F(t_1, t_i) - F(t_0, t_i)]\ N\delta$$

Which, in general, will be non-zero. How do we know that this is the correct way to mark the contract to market ? We simply take the time-t_1 expectation of :

$$[L_{t_i} - F(t_0, t_i)]\, N\delta$$

with respect to the natural forward measure of t_{i+1}

$$E^{P_{t_{i+1}}}_{t1} [Lt_i - F(t_0, t_i)]\, N\delta \;\; = [F(t_1, t_i]\, N\delta = [F(t_1, t_i) - F(t_0, t_i)]\, N\delta$$

where we use the fact that under the $P^{t_{i+1}}$, the $F(t_1,t_i)$ is an unbiased estimate of L_{t_i}.

6.2. Principal CMS Swaps

Pricing CMS swaps is known to involve convexity adjustments. Staying within the context of the simple framework used in this chapter, the industry first obtains the t_1 x t_2 and t_2 x t_3 swaption volatilities. Then, knowing that the swap is a Martingale under the "annuity" measure treated in Chapter 18, various transformations under specific assumptions are performed and then the convexity correction to the forward swap rate is estimated. In other words, the industry calculates the e_t in th equation.

$$cms_t = s^f_t - E_i$$

where cms_t is the CMS rate, s^f_t is the relevant forward swap rate, and e_t is the convexity correction.

It is straightforward to price CMS swaps using the forward Libor dynamics discussed earlier and then use successive measure changes for the required mean corrections. Because CMS swaps offer a good example for such an application, we show a simple case.

Consider a two-period forward CMS swap where a fixed CMS rate xt_o is paid at times t_2 and t_3 against the floating two-period cash swap rate at these times. The present value of the cash flows under the P_{t3} forward probability is given by

$$0 = E^{Pt3}_{t0} \left[(x_{t0} - s_{t1}) \frac{1}{(1+L_{t0}\,\delta)(1+L_{t1}\delta)} + (x_{t0} - s_{t2}) \frac{1}{(1+L_{t0}\,\delta)(1+L_{t1}\delta)(1+L_{t2}\,\delta)} \right] N \qquad (133)$$

where the settlement interval is assumed to be one, and N is the national swap amount. The s_{t1} and s_{t2} are the two-period swap rates unknown at time to. They are given by the usual spot swap formula shown in (50).

Setting δ =1, and rearranging this equation, we obtain.

$$x_{t0} = \frac{E^{Pt3}_{t0}\left[s_{t1} \frac{1}{(1+L_{t0})(1+L_{t1})}\right] + s_{t2} \left[\frac{1}{(1+L_{t0})(1+L_{t1})(1+L_{t2})}\right]}{E^{Pt3}_{t0}\left[st, \frac{1}{(1+Lt_0)(1+L_{t1})}\right] + \left[\frac{1}{(1+L_{t0})(1+L_{t1})(1+L_{t2})}\right]}$$

Hence, to find the value of the CMS rate $x_{t\,o}$, all we need to do is write down the dynamics of the forward Libor processes, $F(t_o, t_1)$ and $F(t_0, t_2)$, under the same forward measure Pt3 as done earlier, and then select Monte Carlo paths.

It is clear that proceeding in this way and obtaining Monte Carlo paths from the arbitrage-free forward Libor dynamics requires calibrating the respective volatilities s^i. But once this is done, and once the correction factors are included in the proper equations, the Monte Carlo paths can be selected in a straightforward manner. The CMS rate can then be calculated from

$$x_{t0} = \frac{\Sigma E^{Pt3}_{t0}\left[s_{t1} \frac{1}{(1+L_{t0})(1+L_{t1})}\right] + s_{t2} \left[\frac{1}{(1+L_{t0})(1+L_{t1})(1+L_{t2})}\right]}{\Sigma E^{Pt3}_{t0}\left[st, \frac{1}{(1+Lt_0)(1+L_{t1})}\right] + \left[\frac{1}{(1+L_{t0})(1+L_{t1})(1+L_{t2})}\right]}$$

where the swap rates s^j_{ti} themselves depend upon the same forward rate trajectories and, hence, can be calculated from the selected paths.

The same exercise can be repeated by starting from perturbed values of volatilities and initial forward rates to obtain the relevant Greeks for risk-management purposes as well.

Notes -

Should Firms Use Derivatives to Manage Risk?

David Fite, Vice President, Bankers Trust

Paul Pfleiderer, Professor, Graduate. School of Business, Stanford University

Introduction

Over the past few decades, the performance of financial markets has been greatly improved by the development of new technologies in communications and information processing. Some argue that financial market performance has also been improved by the creation of new types of securities, especially the financial instruments known as derivatives. The creation and widespread use of derivatives has been brought about in large measure by conceptual advances that have allowed various financial institutions to value and hedge these complex instruments. While the use of derivatives has become widespread, it has also become controversial. Some question the value of derivatives and call for restrictions on their use and new regulations. In this article we consider the economic role played by derivatives. We focus in particular on the use of derivatives by corporations to hedge risks. Should corporations hedge risks; and if so, which ones and why?

Before addressing the issue of corporate hedging, it is useful to consider the general role derivatives can play in financial markets. Financial markets create value in free market economies by performing a number of important functions. One is the allocation of scarce capital to its most productive uses. Should a billion-dollar electric power plant be built, or should the billion dollars be spent instead on the development of a new commercial aircraft? Any elementary textbook on finance or economics shows that financial markets, by establishing the cost of capital for different types of projects, help direct investment to its most productive applications. Financial markets also create value by facilitating an efficient distribution of risks among risk bearers. If the power plant is built and the demand for electricity falls, who should suffer the consequences? Should this be the same party that bears the risk of increases in the price of the coal that fuels the plant? These functions of allocating capital and risks are obviously closely related. In general, the optimal allocation of capital depends on how efficiently risks can be shared among investors. In particular, if the risks created by a given investment can be more efficiently allocated among risk bearers, then the cost of capital for that investment may be reduced and the investment becomes more attractive.

Derivatives have generally lowered the cost and increased the precision with which the market is able to unbundled and distribute risks among risk bearers. However, from this it does not immediately follow that corporations should use derivatives to hedge risks. After all, it might be argued that the ultimate bearers of risk are individuals and not corporations. Corporations can trade risks among themselves, but would such trading have any consequences for how these risks affect individual investors? Since individual investors typically hold positions in many corporations and can themselves alter their risk exposure by trading in derivatives, the case for corporate hedging is not immediately obvious.

We argue below that various "market imperfections" create a solid case for corporate hedging. It is not correct to view corporations as simply passing risks through the corporate structure to individuals who then adjust their positions to attain their optimal risk exposure. Some risks affect corporate earnings in ways that individual investors cannot offset by altering their own financial positions. Because of this, there are several valid reasons for corporations to hedge risks. At the same time, there are some reasons that have been given for corporate hedging that do not make economic sense. We do not intend to provide an exhaustive list of valid reasons for hedging. Undoubtedly many readers can think of specific cases for hedging that do not fall neatly in the taxonomy we present. Our hope is that we have identified most of the major justifications for corporate hedging as well as some of the more dubious ones.

The question as to whether or not a corporation should hedge is not well posed. To have a meaningful discussion, we must specify an objective for the corporation and then ask if and how hedging advances the corporation toward this objective. The standard objective used in a context such as this one is the maximization of shareholder value or shareholder wealth. For most of the following discussion, we assume this is the goal; but in some cases this is problematical-especially where significant differences exist among a firm's shareholders. For example, if some shareholders hold diversified portfolios while others have concentrated holdings in a firm, the two groups will not necessarily agree on the value of a hedging program. In such a case there will not be a dear, unambiguous measure of shareholder value.

Three Views of Corporate Hedging

To frame our discussion of corporate hedging, we begin by examining three frequently encountered views about the value of hedging programs. Our arguments against these lead us to conclude that the only valid justifications for hedging on corporate account are those based on "market imperfections."

Hedging, Since it Reduces Risk, is Good

Financial markets provide a wealth of evidence supporting the notion that investors are risk averse and demand a premium (in terms of a higher expected return) when they hold risky positions.! Hedging, it might be argued, reduces the volatility of a firm's earnings and by doing so makes the cash flow stream delivered by a company to its investors less volatile and therefore more valuable. According to this line of argument, the firm reduces volatility through hedging, which reduces the risk premium investors demand to hold its stock and bonds. This, in turn, raises the value of the future cashflows delivered by the firm since they will be discounted at a lower rate, and this increases the value of the firm and shareholders' wealth.

Hedging, Since it Reduces Rrisk, is Bad

This polar view of hedging is based on the argument that value is generally created by taking on risk and not by avoiding it. A firm that wants to reduce risk can always do so by not investing in risky projects. In the extreme, it can reduce all risk by investing only in short-term government securities. Obviously this course of action produces no value for the shareholders. Thus it is argued that a firm must take on reasonable risks to create value; and if it avoids or transfers these risks, it gives up this value.

Both these arguments are based on incorrect views of how the market "prices" risk. Some risks are priced in the financial markets in the sense that investors require a higher rate of expected return to bear those risks. Other risks are not priced in the sense that investors require no adjustment in expected return. A firm generates value for its investors only when it makes an investment that has a higher expected return than investors require as compensation for the risk.

Hedging is Neither Good nor Bad; it is Irrelevant

Unlike the first two views which rely on simplistic notions of the relations between risk and value, this view is based on a more subtle understanding of the alternatives available to investors and how these investors view various risks.

Part of this argument for the irrelevancy of hedging is based on the fact that many sources of volatility within a firm are not risks that investors care about since these risks nearly vanish in a diversified portfolio.2 In particular, an investor who holds a diversified portfolio is not affected in any significant way by sources of volatility that affect only one or a few firms. Such "idiosyncratic" risks are inconsequential for investors. Consider a well-diversified investor who holds stock in Ford. Even though Ford is a large firm with a market capitalization of $18.854 billion (as of December 1992), its capitalization is only 0.47% of the total value of NYSE stocks and an even smaller fraction of the total value of all U. S. stocks.

Assume there is a risk that affects only Ford and that this risk either adds 10% to Ford's return or subtracts 10% from it. Then a diversified investor who holds Ford in same proportion as its value in the NYSE stock portfolio will see his return vary by at most 0.047% due to this risk. To put this into perspective, note that a diversified investor with a total investment of $100,000 loses only $4.70 as a result of a 10% decline in Ford stock. Such an investor would see little value in Ford's removing this risk and would oppose its removal if there were a significant cost involved. The diversified investor is only concerned with pervasive risks, i.e., risks that affect a large number of firms. The idiosyncratic risk at Ford has virtually no effect on any investor who has allocated his wealth across many firms.

Even when we consider pervasive risks, it is not immediately dear that a firm can gain by hedging these risks. Oil price shocks are good examples of pervasive risks that affect many firms. Note, however, that unexpected changes in the price of oil affect firms in different ways.

An increase in the price of oil will generally increase the earnings of the oil companies but will decrease those of the airlines. A diversified investor who holds both oil and airline stocks is therefore at least partially hedged in his portfolio against oil price shocks. Such a diversified investor's risk exposure is essentially unchanged if the airlines take long positions in oil futures contracts and obtain them from the oil companies who take the offsetting short sides. Of course, there are pervasive risks that affect most firms in the same way. These risks do not "cancel out" in a diversified portfolio, and it is reasonable to assume that diversified investors will be sensitive to them. Does this mean that a firm is better off if it reduces these risks? Not necessarily.

Assume that an unexpected increase in energy costs affects almost all firms adversely. This, then, is a pervasive risk that does not cancel out in diversified portfolios.[3] Assume that because energy cost risk cannot be diversified away, diversified investors require a higher expected return on securities with a high sensitivity to energy costs than they do for securities with lower sensitivity. If a firm could change its real operations in some costless way that reduced its exposure to energy costs, then its value would increase. However, if that firm reduced its energy cost exposure by hedging, then it would pass energy risk on to some counterparty who would demand compensation for bearing this risk. In an efficient market, that compensation is precisely equal to the increase in the value of the firm's cashflows due to the energy risk reduction. The shareholders will be neither better nor worse off. Thus it is irrelevant what the firm does to "manage" this risk.

Even if one is unwilling to accept the arguments made above for the irrelevancy of corporate hedging policies, one still must contend with another argument for irrelevancy, this one based on the well-known and important insights contained in Modigliani and Miller's analysis of a firm's capital structure.. Modigliani and Miller considered changes in a firm's financing policy that do not alter the firm's investment policy. They showed that such changes neither increase nor decrease the firm's value when there are no transactions costs, no taxes, and no information asymmetries. Their argument was based on the observation that any financial position that the firm can achieve by altering the set of claims it issues can also be achieved by holders of the firm's debt and equity if they adjust their "own account" positions. At the same time, any investors unhappy with the change can undo it by trading on their own accounts.

In their original analysis, Modigliani and Miller focused on the corporation's debt/equity ratio (i.e., its use of leverage), but their arguments clearly apply with equal force to a firm's hedging strategy. Under the assumptions made by Modigliani and Miller, there is no reason for a firm to hedge since investors can do it on their own accounts if this is something they desire. For example, a U. S. export firm that sells in Germany can reduce its exposure to exchange risk by taking a position in the $/ DM forward market. However, any of the firm's shareholders who value this risk reduction can achieve the same result without the firm's hedging by taking a similar (but smaller) position in the forward market on their own account. However, some of the shareholders might actually prefer the export firm's exposure to exchange risk - perhaps because they are importers of German goods and their risk is offset to some extent by the exporter's position. If the exporter does hedge its exposure, these shareholders can undo the result by taking the opposite and offsetting position on their own accounts. In all cases, it makes no difference what the exporter does as long as the shareholders know the exporter's hedging positions and can trade in the same instruments.

The General Conditions for Corporate Hedging to Benefit Shareholders

If value is created by altering risk exposure through derivatives, it must be because one or more of the assumptions used by Modigliani and Miller do not hold. It is clear that at least two conditions must be met by any worthwhile corporate hedging strategy: (1) It must change the firm's cash flows in a way that shareholders' value and the benefit to shareholders must be greater than the cost of hedging; and (2) Hedging on corporate account must be the least expensive way to bring about the beneficial change in cashflows. In particular, the firm must be at least as efficient in adjusting the risk exposure and creating the improvement as shareholders would be if they hedged on their own account.

One might think that the second condition hardy merits serious attention. Doesn't the firm always have a cost advantage over individual shareholders since it can take advantage of scale economies and therefore pay lower transactions costs? In fact, there are many situations where shareholders have a distinct advantage over firms in controlling risk exposure. For example, in the 1960's, a large member of conglomerate mergers occurred in the U. S. At the time, one of the justifications given for these mergers was that by creating a diversified company, a conglomerate merger lowered the risk experienced by shareholders. While there may have been good reasons for the formation of conglomerates, this almost certainly was not one of them. Individual shareholders could on their own accounts achieve the benefits of diversification simply by buying shares in a number of companies in different industries. 'This could be accomplished at far lower cost (especially if it were done through mutual funds) than the cost of "physically" merging several companies. Corporations are high-cost producers of diversification, at least when they do it through mergers. It is therefore reasonable for us to ask whether the second condition is met by corporate hedging strategies. As we show below, for many of the benefits produced by hedging, the corporation is likely to be the lowest-cost producer. In fact, we argue that several benefits of hedging cannot be produced at all by the shareholders' hedging on their own accounts and that only corporate hedging can bring about these gains.

Since we measure the value of corporate hedging in terms of the effects hedging has on shareholder wealth, we must now consider how shareholders value the cashflows produced by a firm. Many approaches break the problem of valuing a firm's shares into two parts. In the first step, future cashflows that will be produced by the firm and paid to shareholders are forecasted. In the second step, these expected cashflows are discounted to the present using the appropriate risk adjusted discount rate. This discount rate is the rate of return shareholders require as compensation for the riskiness of the firm's cashflows. It follows that hedging can affect the firm's value by changing the expectation of its future cashflows, by changing the discount rate shareholders use to discount these cashflows, or by doing both. Since corporate hedging polices alter the firm's risk exposure, it would seem that the most profound effects of hedging would be felt through changes in the firm's risk-adjusted discount rate.

Paradoxically, we argue below that most of the gains produced by corporate hedging for shareholders are due to increases in expected cashflows and not to reductions in the discount rate. Along these lines, we divide the remaining discussion into two parts. First we discuss the effects of corporate hedging on the risks shareholders actually bear. This relates mainly to the discount rate. Then we discuss how risk actually affects the firm and more specifically how corporate hedging affects the expected cashflows the firm can generate and deliver to shareholders.

Risks Shareholders Bear and Benefits of Corporate Hedging

Does corporate hedging produce any gain to shareholders by changing the risk they experience? Much of the foregoing discussion leads one to be skeptical that it does. Since shareholders tend to hold well diversified portfolios through intermediaries such as mutual funds and pension funds, many corporate risks are diversified away in the typical shareholder's portfolio. In addition, the Modigliani-Miller argument reminds us that shareholders have the ability to control their risk exposure on their own account and do not necessarily require that corporations do it for them. Do these arguments for the irrelevance of corporate hedging hold up after we explicitly consider transactions costs and other market imperfections?

First, note that not all shareholders hold well-diversified portfolios. Many U. S. corporations have shareholders who own fairly large stakes in the firm and for whom this stake is a large portion of their net worth. In many cases, these are founders or others who for control reasons have acquired a significant number of shares. These undiversified shareholders will quite likely differ from diversified shareholders in the way they view the risks faced by the corporation. We have already noted that in this case it is somewhat problematic to talk about the affects of hedging on shareholder value. Shareholders will not value corporate hedging in the same way.

For example, assume that a firm has one large shareholder who holds 30% of the firm's shares and that this stake represents almost all of his wealth. The remaining shares (70%) are held by diversified investors. Now assume that the firm faces foreign exchange risk since it is a net importer from Canada, but also assume that the diversified shareholders on net have no foreign exchange exposure. This can come about if the diversified shareholders hold stock in both exporters to and importers from Canada. In terms of their risk exposure, the diversified shareholders will not benefit if the corporation hedges the exchange rate risk, but clearly the undiversified shareholder will see his risk reduced. Should the firm hedge?

Assume that the firm can costlessly eliminate the foreign exchange risk by trading a futures contract with a firm that exports to Canada but that the cost of the undiversified shareholder's hedging on his own account is greater than zero. An argument can be made for the firm to hedge since the diversified shareholders are no worse off and the large shareholder is better off.[5] Now assume that it is somewhat costly for the firm to hedge the exchange rate risk due to transactions costs, but that this cost is less than the cost the undiversified shareholder pays if he hedges on his own account Should the firm hedge? If it does, the diversified shareholders are worse off since they pay 1'0% of the cost of hedging but receive no gain. The diversified shareholder potentially gains, however, since he only pays of the costs of hedging but benefits from the lower risk exposure. Given the assumptions we have made, it is clear that if we put the matter to a shareholder vote and everyone votes in his or her own interest, the vote will go against hedging.

Before concluding that hedging will always be rejected by shareholders when the majority of shareholders are diversified, we should ask if the large shareholder's presence in any way benefits the diversified shareholders. Several arguments have been made to support the notion that the presence of a large shareholder does benefit the other shareholders. Many of these are based on "free-rider" problems that occur when shareholding is widely diversified. It is clear that when each shareholder holds a very small stake in each firm, none has much of an incentive to monitor the performance of the firm or to pressure its management into making value-improving changes. A diversified shareholder who pays the substantial costs of time and effort involved in monitoring and lobbying for changes receives only a tiny fraction of the gain. He prefers that others pay these costs and that he "free ride" on their efforts. Of course, the free-rider problem exists even when there is a shareholder who holds a stake of, say, 30%, but in this case it is not as severe. A shareholder with a 30% stake has some incentive to monitor since he receives 30% of the gains rather than the minuscule part a diversified investor receives. In fact, the larger the stake, the greater the incentive to monitor.

Since this monitoring produces a gain for all of the shareholders, diversified shareholders may want to encourage a few shareholders to maintain large positions in the firm. One way to do this is to lower the costs paid by those investors who take large stakes. Obviously one of the major costs borne by these investors is the added risk exposure due to the loss of diversification. A corporate hedging program reduces this cost and encourages the undiversified investors to maintain larger stakes than they otherwise would ultimately this increase expected cashflows since with more concentrated ownership, more monitoring occurs.[6]

Now we turn to the diversified investors' exposure to risks. Again, diversified investors include a large number of securities in their portfolios and put a small weight on each. Although investors who hold well diversified portfolios are hardly affected by risks that are felt by only one or two firms, they generally have reason to be concerned with risks that are pervasive, i.e., risks that are felt by many firms. Pervasive risks typically do not vanish in diversified portfolios.[7] Diversified investors will therefore prefer that firms hedge pervasive risks rather than non pervasive risks.[8] It must be emphasized, however, that diversified investors as a class gain only if the pervasive risks are transferred "out of the system." This means that when corporation A hedges a risk, it is not absorbed by corporation B, a company in which the diversified shareholders also hold stock. If risk is not transferred out of the system, the risk exposure of the average diversified investor remains the same. If, however, corporation A hedges the risk by transferring it to a privately held company or to a foreign company in which the diversified investors do not hold shares, then the average diversified shareholder may gain.

Hedging that transfers risk "out of the system" is one way to expand the set of securities over which investors diversify. Consider international diversification. The gains to international diversification appear to be quite large, yet most investors continue to concentrate their portfolio holdings in their domestic markets. Under certain circumstances, hedging can provide a way for these investors to realize some (but by no means all) of the gains of international diversification. For example, if the average firm in the country of Sell petrol has positive exposure to the risk of changes in oil prices (returns increase when oil prices rise) while the average firm in the country of Buy petrol has negative exposure, the inhabitants of both countries can reduce the variance of their portfolios' returns by buying diversified portfolios consisting of shares in Sell petrol and Buy petrol. If, for some reason, the inhabitants of each country do not diversify in this fashion but instead hold only portfolios diversified over their domestic stocks, then risk can still be reduced if the companies in each country hedge their risk of oil price exposure. This might be done by having the companies in Sell petrol take short positions in oil futures and the companies in Buy petrol take the offsetting long positions. It could also be done if the investors of Sell petrol issue short futures contracts to the investors of Buy petrol.

Given resistance or impediments to international diversification, hedging is potentially valuable to the shareholders of each country; but at this point there is no reason for it to be done at the corporate level. The obvious justification for corporate-level hedging is the savings in transactions costs. The potential sources of these savings are obvious. If fixed costs are associated with a hedge, it is better for a firm to pay these costs once on behalf of all shareholders than for each shareholder to pay these costs individually. A major component of these fixed costs is the cost of acquiring the information about the firm's risk exposure. Clearly there are also possibilities of reducing trading costs, legal costs, and so on, when hedging is done at the corporate level.

It would seem that the transactions cost advantages that firms possess would always decide in favor of hedging on corporate account. Surprisingly, this is not always true. The question of who should hedge the corporation or shareholders - becomes much more complicated once we acknowledge that not all diversified investors are alike. This is because diversified investors are exposed to risks that affect them outside of their investment portfolios.

An airline pilot, for example, is exposed to oil price risk by virtue of his occupation.9 If the pilot is a savvy investor, he skews his investment portfolio away from stocks that are negatively affected by oil price risk (i.e., have low returns when oil prices increase) and toward those that are positively affected. This reduces his overall exposure to oil price risk. If the pilot holds a stock in a firm that has a negative exposure to oil price risks, he prefers that the firm hedge the risk if it is not too costly to do so. At the same time, an Exxon employee who holds stock in a company with a negative exposure to oil price risk prefers that the company bear this risk. The negative exposure to oil price risk is a valuable hedge for the Exxon employee since it tends to offset his positive exposure to oil price risk resulting from his employment in the oil industry. Thus even shareholders who hold diversified portfolios may disagree over what a particular firm's hedging policy should be.

This disagreement among shareholders may mean that the best policy for the firm is not to hedge even if a majority of its shareholders benefit and if the firm's transactions costs are lower than those of its shareholders. For example, consider a firm with 100 shares and assume that it has a negative exposure to oil prices. Suppose that 60 of its shares are held by airline pilots and the remaining 40 by oil company employees to finally, assume that it costs twice as much for shareholders to hedge a unit exposure to oil-price risk as it does for the company to hedge the same unit of exposure. If the company hedges its negative exposure to oil price risk, it pays 100c, where c is the cost the company pays per share to hedge. The oil company employees are, however, worse off than they were before since they have lost the risk reduction produced by the company's negative exposure to oil price risk. As a consequence, they must unwind the hedge on their own accounts. This costs them 40(2c) = 80c. (Recall that individual shareholders pay twice the transactions costs that the firm pays). Thus the total transactions cost spent when the firm hedges is 180c. If the firm does' not hedge, then the airline pilots must hedge on their own. This will cost 60(2c) = 120c, which is less than l80c.

Of course this is only a partial analysis of the problem. If the company does hedge, then it becomes more attractive to airline pilots and less attractive to oil company employees. The mix of shareholders may change from the 60/40 mix we assumed. The point remains that if shareholders have differing risk exposures due to such factors as their occupations (their human capital) and their undiversified real estate holdings, they will have conflicting preferences about corporate hedging programs. Hedging to meet some of the shareholders' needs may be worse than not hedging at all.

This brings us to the final consideration concerning shareholder risk exposure. Since shareholders have differing risk exposures outside of their security market portfolios (again consider the pilot and the oil company worker), they form "clienteles" for various stocks that serve as good hedges for these no market risks. These shareholders want the risk exposure of the stocks they are buying to remain relatively constant over time. If the risk exposures of firms' shares were subject to frequent and major changes, then investors would need to closely follow the firms in which they invest. If exposures were changing significantly, these investors would often find it necessary to trade their shares to reestablish their optimal positions. This places a burden on the shareholders that can be avoided if companies follow a hedging policy that keeps the exposure of their shares relatively constant even if their operational exposure to risks is changing. This argument for stabilizing a stock's risk exposure in the interest of a clientele resembles arguments made concerning dividend policy. It is suggested that some investors desire dividends and form a natural clientele for high-yield stocks; others prefer "growth" and seek low-yield stocks. Firms do not necessarily gain by following a high-yield or low-yield strategy. What is important is that the firms not vary its payout significantly quarter to quarter or even year to year.

We have argued that there are justifiable reasons for corporate hedging based on its effects on shareholder risk bearing. However, in all of these cases, the gain is probably modest (at least for investors that hold well diversified portfolios). The biggest gains may come when hedging substitutes (partially) for international diversification or allows shareholders to share risk with privately held firms or other firms in which diversified shareholders cannot trade. The issue is complicated by the fact that there are cases where a hedging program might benefit some shareholders but make others worse off.

Even if the gains created by corporate hedging and the lowering of shareholders' risk are typically small, this does not mean that hedging is not worthwhile. After all, the costs of hedging are often also small. We could attempt to quantify these gains and costs in particular situations, but this would not be easy - nor is it necessary. Most of the value of corporate hedging is not due to how it alters the risks experienced by shareholders but rather to how it alters the risks experienced by the firm itself. We now turn to this issue.

The Effects of Risk on the Finn and the Benefits of Corporate Hedging

In considering the effects of risk on the firm itself, we do not want to fall into the trap of assuming the firm should be treated as a separate individual with preferences of its own. We recognize that the firm is owned by the shareholders and that the firm's behavior is determined by the interaction of a number of individuals who may have conflicting interests. It has been argued that since the firm is not an individual with preferences, it is inappropriate to characterize the firm as risk averse. In some cases, this has been interpreted to mean that the firm should be considered "risk neutral" Discussions along these lines are generally not very fruitful and are often misleading. Nevertheless, we argue that often the firm should, from the perspective of its shareholders, behave as if it is risk averse. This risk aversion on the firm level creates the demand for corporate hedging and risk management.

Individuals are risk averse if they value the gain of any given dollar amount less than they value the 10ssDf the same dollar amount. A risk-averse individual rejects a gamble that gives an equal chance of winning and losing $10,000; the 50% chance of having an extra $10,000 does not make up for the 50% chance of having $10,000 less. Now consider a firm that accepts the following gamble: with 50% probability, its earnings (before interest and taxes) will be $20 million higher than otherwise;

and with 50% probability they will be $20 million lower. Does this gamble increase, decrease or leave unchanged the expected amount the firm can deliver to its investors? Suppose that the increase of $20 million result in only $13 million in additional cash for the investors but the investors feel the full effect of the $2o-million loss; when the company loses $20 million, the investors receive $20 million less cash than they would have received if the company had not gambled. In this case, the investors are dearly worse off with the gamble than they are without it.

We emphasize that the investors are worse off not because this gamble added risk to their portfolios. The gamble could be decided by the flip of a coin, in which case it would be purely diversifiable risk. A diversified investor essentially would not care about this risk as it affects the risk of his portfolio holdings. Rather, the investors lose because the gamble reduces the expected amount of cash the firm delivers to them. From an investor's perspective, the value of the extra $20 million in corporate earnings measured in terms of the extra cash delivered to the investors is less than the value of the $20 million loss in corporate earnings measured in terms of cash lost to the investors. This means that even if an investor were risk neutral, he would want the firm to behave as if it were risk averse when it considered this gamble.

This raises the key question: do a firm's investors lose more when a dollar of earnings is lost than they gain when an extra dollar is earned? Note that this asymmetry does not occur when investors buy shares in an open-end mutual fund. Within a mutual fund, gains and losses are generally symmetric. A dollar earned is one more dollar available to the fund's investors, and a dollar lost is one less dollar. The amount the mutual fund can distribute to its investors is a linear function of the amount it earns.[11] If there are asymmetries in gains and losses within a corporation, it will be because of some nonlinear relation between earnings and what investors receive. We now explore some of these, starting with taxes.

Corporate Taxation and Hedging

Consider a firm with a corporate tax rate of 40%. Assume that if this firm does not hedge any of its operating risks, it will each year either earn $250 million with probability 75% or lose $50 million with probability 25%. Given these probabilities, the firm's expected earnings each year are $175 million. Assume that the company has the opportunity to fully and costlessly hedge its risks away. This means that the company will receive $175 million per year for certain. Should this company hedge?

Assume that it does not. Then when the firm earns $250 million, it will have $150 million after taxes to distribute to shareholders. (To simplify matters, we assume that the company has no debt so that all after-tax earnings are paid to shareholders.) When the company loses $50 million, it pays no taxes. If we assume that it can carry these losses forward (with interest) and use them fully to offset future taxes, then the loss to shareholders is not $50 million but only $30 million since future tax liability is reduced by $20 million. Thus the expected cash available to shareholders is 0.25 x (-$30,000,000) + 0.75 x ($150,000,000) = $105,000,000. If the company hedges, the cash available is also $105,000,000 since this is 60% of $175,000,000. Hedging has not changed the expected amount of cash available to shareholders.

Hedging did not have any effect because we assumed a uniform tax structure which treats losses and gains symmetrically by allowing the firm to take full advantage of losses carried forward or backward. In actuality, most tax structures are not linear. Tax rates often rise as income increases, and corporations cannot fully realize the tax benefits of losses as we assumed. This creates a role for hedging. Assume that when a $50 million loss is incurred, future tax liability is only reduced by $10 million, not by $20 million as we assumed above. This could be due to limitations on the ability to carry losses forward or backward, reductions in value of these offsets due to the time value of money, and 50 on. If the firm does not hedge, the expected amount available to shareholders is only 0.25 x ($40,000,000) + 0.75 x ($150,000,000) = $102,500,000. This is $2.5 million less than the amount available to shareholders when the firm hedges. Even if hedging is costly, as long as the cost is under $2.5 million, shareholders are better off with corporate hedging. Hedging is valuable here because with the asymmetric tax structure, shareholders lose more when the company's before-tax income falls by a given amount than they gain when it rises by an equal amount. In a sense, the tax structure makes the corporation risk averse.[12]

Finally and importantly, this gain can only be produced by hedging on the corporate level; shareholders cannot hedge on their own accounts and reduce the corporation's tax liability in the manner shown above.[13]

Hedging and the Costs of Bankruptcy and Financial Distress

Any good corporate finance textbook has a long disquisition on the costs of bankruptcy and financial distress. These costs are usually cited as one of the reasons why firms do not fully exploit the tax advantages of increasing leverage. Since these costs are described in such detail elsewhere (Brealey and Myers 1991; Ross, Westerfield and Jaffe 1993; and Van Home 1992), we summarize them briefly here and then discuss the obvious role hedging plays in reducing these costs. When a levered firm defaults on its debt or enters bankruptcy proceedings, direct costs are incurred through the increased need for legal, accounting and other professional services. While these direct costs are not necessarily trivial, it is usually claimed that the indirect costs of financial distress and bankruptcy are the most significant. These indirect costs take many forms.

For example, in situations of financial distress, the attention of upper management may be diverted from managing the firm's operations. This generally results in a loss of value. Due to uncertainties in how bankruptcy proceedings will be resolved, customers may be more reluctant to buy and suppliers may be more reluctant to make costly supply commitments when the value of these transactions depends on how long and in what form the firm remains in business.[14] Conflicts among various claimholders may cause the firm to pass up profitable investment opportunities. These conflicts occur when a firm is near bankruptcy and any new investment by shareholders will mainly benefit the bondholders. The simple solution to this problem is to reorganize the firm in such a way that the conflict no longer exists and then raise the funds necessary to undertake the profitable investment. In practice, such reorganization takes time and may not be achievable. These are just a few examples of indirect costs.

While some of the costs of bankruptcy and financial distress are subtle, the role hedging can play in reducing these costs is obvious. Hedging generally lowers the probability of financial distress and bankruptcy. By lowering the probability, hedging lowers the expected costs of distress and increases the expected cashflows available for shareholders. Again, the gain produced by hedging is due to an asymmetry. If earnings are low or negative, the shareholders must pay the costs of financial distress and perhaps bankruptcy.[15] If earnings are high, the shareholders do not get any extra bonus (e.g., a reverse payment from bankruptcy lawyers) to make up for the costs on the downside. Finally, note that when hedging reduces the cost of financial distress, it also increases debt capacity. Thus the gain due to hedging may show up through the firm's ability to increase its degree of leverage and realize the tax advantages or other benefits of a higher debt-to-equity ratio.[16]

Hedging and the Cost of Funding New Investment

The simple rule often given for choosing investment projects is the net present value rule: choose those and only those projects with positive net present values (NPVs). A project has a positive NPV if the present value of the cashflows it produces is greater than the investment required to undertake the project. Value is left lying on the table whenever a positive net present value project is not undertaken.[17] Unfortunately, a firm's current shareholders may find that they are better off passing up a positive NPV project. This occurs when the project cannot be funded out of retained earnings and outside financing is required.[18] Consider a firm that has 100 shareholders, each of which owns one share. The firm's management knows that the total value of the firm's assets is $1,000,000. Thus each shareholder currently has a claim worth $10,000. Now assume that the firm can undertake an investment project which costs $500,000 but which will produce cashflows worth $700,(XX). The net present value of this project is therefore $200,000. The firm, however, has no retained earnings to fund the project, so it must issue new shares to raise the capital. The firm's problem is that the outside market only values the assets the firm has in place at $500,000, not $1,000,000.

What could give rise to this discrepancy in valuations? Those who firmly believe in efficient markets would probably conclude that the management is mistaken and that it overestimates the value of the firm by $500,000. This, of course, is possible. But it is also possible that the management has information about the value of assets in place that the market does not have. For example, the company might be a biotech firm which is developing an experimental drug. The management could have some information relating to the prospects for success that cannot be quantified and is not subject to disclosure requirements. Indeed, the management may have compelling strategic reasons to keep the information secret. Thus, even if the management could voluntarily disclose the information, it might find it too costly to do so.

Assume that the management's valuation of $1,000,000 is correct, and consider what happens if the management raises capital to undertake the project. To do this, an additional 100 shares must be issued to raise $500,000. This means that each of the original shareholders will own 1 / 200 of the firm. The true value of the firm will be $1,000,000 + $700,000 = $1,700,000. Each of the original shareholder's stakes thus falls in value to $8,500. Recall that if the project is not undertaken, the value' of each stake is $10,000. Quite simply, the dilution that occurs when shares are issued at prices below their true value overwhelms the increase in value brought about by the positive NPV project. The shareholders are forced to leave money on the table.

This problem would not have occurred had the firm possessed sufficient internal funds to undertake the project without raising capital on the outside. For example, if the firm had $500,000 in retained earnings in addition to its $1,000,000 in fixed assets, it could undertake the project and increase its value from $1,500,000 to $1,700,000. The original shareholders would not be hurt by dilution and would capture the full $200,000 increase created by the positive NPV project. All of this points to another valid reason for hedging on corporate account. Consider a firm that will over the years have a sequence of valuable investment projects to undertake, and assume that in a typical year it will have sufficient internal funds to finance the projects available that year. However, in those years when earnings fall to very low levels, the firm will not have sufficient internal funds to undertake positive NPV projects and may find itself in the predicament described above. A hedging strategy that stabilizes earnings and lowers the likelihood of the firm's needing outside capital is valuable since it reduces the chance that profitable investment projects will be foregone. In fact, any additional cost associated with outside financing (underwriters' fees, market price impact, etc.) creates a rationale for stabilizing earnings through hedging.[19]

Hedging and Agency Costs

We have shown that in many circumstances, reducing the volatility of earnings increases the expected amount of cash shareholders will receive. One should not conclude from this that shareholders always desire lower volatility. In fact, when the firm has substantial leverage (i.e., a high debt-to-equity ratio), shareholders have strong incentives to increase volatility. Additional risk or volatility tends to raise the value of the shareholders' position in a levered firm whenever the shareholders receive the benefits of the "upside" while the debtholders suffer the consequences of the "downside."

The possibility of the shareholder taking advantage of the debtholders' with a "heads I win, tails you lose" gamble is one of the sources of what has come to be termed "the agency costs of debt." The following example illustrates these agency costs and shows how hedging might be used to reduce these costs. Assume a find has a single debt liability of $700 million which is due in one year. Suppose that if the firm continues to operate in its current manner, it will have assets worth $600 million when the debt comes due. This means that the find will be bankrupt, shareholders will receive nothing, and debtholders will receive only $600 million of the $700 million owed to them.

Now assume that the company can change its operations and follow a risky strategy. If the risky strategy pays off, the firm will be worth $800 million; if the strategy fails, the find will only be worth $400 million. If these two outcomes (success and failure) are equally likely, the expected payout to shareholders will be $50 million (50% chance of receiving $100 million - which is the residual from the $800 million once the $700 million in debt is paid - and 50% chance of receiving zero). The expected payout to bondholders is $5SO million (50% chance at $400 million and 50% at $700 million). The risky strategy has not increased the expected value of the firm's assets (this remains $600 million), but it has transferred $50 million in expected payout from the bondholders to the shareholders. Of course, the bondholders are well aware of this possibility when they purchase the bonds and use bond covenants to restrict shareholders from following risky strategies.

Assume that the shareholders are prevented by covenants from taking a risky strategy of the sort described above. Is this a problem? Consider again the above example but with one change: assume that if the risky strategy pays off, the value of the firm is $1,200 million, not $800 million. The risky strategy has now increased the expected value of the firm from $600 million to $800 million (the average of $400 million and $1,200 million). This strategy is clearly worth pursuing.[20] However, unless the terms of the debt are renegotiated, the debtholders will not favor the strategy and will be unwilling to waive the covenants even though doing so would increase the firm's value. This is because the debtholders continue to have a claim that pays either $400 million or $700 million with equal probability. Since the expectation of this claim is $550 million, they will prefer that the company do nothing since this gives them $600 million for sure. The problem is solved if the terms of the debt contract can be easily renegotiated; but in many cases, especially those of publicly placed debt, this may be costly or impossible. Can hedging solve this problem?

Assume that the risk of the proposed risky strategy can be hedged away. For example, it may be that the risky strategy involves the firm's selling in a foreign market and that much of the risk is due to foreign exchange uncertainties. Assume that w_ 'n this risk is hedged away, the SO/SO gamble of $400 million or $1,200 million becomes $800 million with certainty. Then, without the debt being renegotiated, the debtholders will receive $700 million instead of $600 million, and the shareholders will receive $100 million instead of nothing. By hedging the risk, the firm captures the value of the risky strategy; if the risk had not been hedged and the debt could not be renegotiated, the bondholders would have blocked the firm from obtaining the increase in value.

The example is admittedly simplistic, but the point it illustrates carries over to more realistic and complicated settings. When the capital structure includes debt, shareholders and debtholders may take opposite positions as to the firm's operations since such matters affect the riskiness of the firm's value. Hedging allows risks to be controlled and thus gives the shareholders more flexibility in altering the firm's operations without substantially changing the firm's overall risk. As we have shown, this added flexibility may mean that the firm can make value-improving changes in the way the firm operates - changes which otherwise would have been blocked by the bondholders.[21]

Hedging, Incentives and Employee Compensation

In most cases, the compensation of employees is positively related to the performance of the firms that employ them. If a firm does well, its employees generally receive higher levels of compensation than if the firm does badly. There are at least three reasons that justify this positive relation: risk sharing, constraints on the firm, and incentives.

First, we consider risk sharing. A small shopkeeper with a single employee would probably find it advantageous to pay the employee more when business is good and less when it is bad. This is because the shop owner absorbs all of the risk if the employee is paid a wage that is independent of the level of business in the shop. Unless the shop owner is risk neutral, it is generally better for the owner and the employee to share the risk. This means that on average, the employee must be paid more since the employee must be compensated for bearing some of the risk. However, a risk-averse shop owner will gladly pay a little more to the employee (on average) for bearing some risk since this reduces the shop owner's risk. The employee's variable compensation is basically a hedge for the shop owner.

While risk sharing along these lines makes sense in a small business, it is a less compelling reason for the variable compensation of employees in large corporations with diversified shareholders. The risk of a large corporation is shared extensively among its shareholders and other financial claimholders; there is minuscule advantage in employees' bearing a portion of the risk.[22] In fact, if risk sharing is the only consideration, a substantial loss occurs when employees bear significant risk since they must be compensated for it through higher average compensation. This increased cost is worth much less than the meager benefit the shareholders receive when risk is shifted to the employees.[23]

This brings us to the second reason employee compensation might vary with the firm's fortunes: in bad times, the firm may be constrained to pay employees less because of market imperfections. Consider a firm that has a wage bill of $50 million and revenues which vary between $25 and $100 million. We have argued that from a risk-sharing point of view, it is generally optimal for the shareholders to absorb most of the risk of variations in revenues or earnings. If the firm has several bad years of revenues at the $25-million level, the shareholders should contribute to make up the shortfall between the wage bill and revenue. If the money is not available in retained earnings, then the firm should raise more capital. But this may be excessively costly if it is even possible. Recall our discussion about hedging and the cost of funding new investment where the issuance of new shares involved substantial dilution. In such a situation, the company may cut back on employee compensation rather than raise funds externally. The employees are forced to bear a risk created by the company's funding constraints. Obviously, hedging can playa role here. If the risk of the revenue stream can be reduced, the company is less likely to have to reduce employee compensation. This means that employees will have more stable incomes and will not require additional compensation for risk. This savings in the wage bill accrues to the shareholders.

The third and final reason for employee compensation to be tied to the firm's performance concerns incentives, especially those for upper management. Over the last two decades, economists have extensively studied incentive contracting issues. This research considers the problems faced by a principal who hires an agent to act on the principal's behalf. It is generally assumed that the principal cannot observe all the agent's actions and in particular cannot observe the agent's level of effort. The optimal incentive contract for a principal to offer an agent can be quite complicated. Among other things, it depends on what the principal can observe, how the agent can affect the principal's welfare, and what degree of risk the principal and the agent can tolerate. In the context of our discussion, the principals are the shareholders of a firm and the agents are the firm's managers. As we have pointed out above, the shareholders are generally well-diversified investors and are much better able than the employees to bear the firm's risks. A number of results in the incentive contracting research concern cases where the principal is risk neutral (or nearly so) and the agent is risk averse. In these cases, the optimal incentive contract for the agent does not expose the agent to a risk unless it creates an incentive for the agent to work harder.

For example, assume that in January, a U. S. company sends an employee to negotiate a one-year supply contract with a French company. Assume that the contract will specify the quantity to be delivered each month and that the monthly payment will be denominated in French francs and fixed up front. Clearly, it is not sensible for the company to pay the employee a bonus in December that is inversely related to the dollar cost of the goods purchased over the year. If the French franc unexpectedly appreciates relative to the dollar over the year, the dollar cost of the good will increase; but it is not sensible to penalize the employee for this since the exchange rate is completely outside of his control. Of course, if the French franc depreciates instead, the dollar cost falls and the employee is rewarded. But again there is no reason for this since the gain was due to an exchange rate change and not to the employee's efforts. Exposing the employee to the risk of exchange rate movements that occur after the contract is negotiated serves no purpose at all in motivating the employee at the beginning of the year to negotiate a better price in French francs.

It would seem that these incentive contracting considerations provide another rationale for hedging on the corporate level. The compensation of the upper-level managers of a corporation is typically tied to various measures of corporate performance such as earnings and stock price appreciation. Stock options, for example, provide an obvious incentive for managers to increase shareholder value since many things affecting the stock price are under the managers' control. However, for almost every company, many determinants of the stock price are beyond the managers' control. If the company is a multinational corporation, unexpected changes in exchange rates can affect the company's profitability and its stock price; but just as in our example above, these typically fall outside the control of managers. It would seem that corporate hedging, since it remove some of these risks, makes stock options more effective in motivating the manager. If exchange rate fluctuations, oil price changes and similar risks are hedged, then changes in the stock price are less likely to arise from factors not under management control and more likely to result from actions taken by tI1e management.

There is a problem with this incentive-based argument for corporate hedging. It provides a reason for hedging certain risks insofar as they affect the amount paid to managers, but it provides no reason for hedging to be done for the entire firm. Assume a multinational firm faces exchange rate risk beyond the control of management. One way to establish the appropriate incentives for managers is to base their compensation on the firm's future stock price performance and then hedge the exchange rate risk for the entire company. Call this Plan A. The same effect, however, can be achieved by Plan B. Under Plan B, the company does not hedge the exchange rate risk but instead adjusts the manager's compensation to remove the effects of unexpected exchange rate movements. Doing this involves determining what the stock price would have been had the company hedged and what the manager's compensation would have been had this been the stock price and had the company adopted Plan A. The manager could then receive this amount. The company would not need to hedge its entire risk to remove this risk from the managers' compensation. (Note that Plan B involves no trading at all by the firm in outside markets. All hedging is done internally by adjusting accounts.)

One could argue that the compensation committee of the board of directors would not have all of the information needed to make these adjustments. The risk exposure of the company might frequently change, and at any time of the year the managers would be the best informed about the need for hedging. It could be argued that managers should be given the opportunity to take the appropriate hedging positions on corporate account throughout the year as opposed to letting a committee guess an appropriate year-end adjustment. While this might seem to justify using Plan A and hedging for the entire corporation, it does not. Plan B can still be implemented if during the 0year the managers report daily the hedging they would do under Plan A. These reports can then be used at the end of the year to make the appropriate adjustments.

Is there any reason to adopt Plan A over Plan B? One possible justification for the use of corporate-wide hedging over hedging only for the managers is more "political" than economic. Consider what might happen if plan B is used and the company experiences a large loss due to a risk beyond the managers' control. Even though the shareholders suffer this major loss, the compensation required for the managers under Plan B might be quite high. This would be true if the managers had performed quite well in terms of those things under their control. In other words, losses would have been even higher had the managers not performed so well. In such a circumstance, managerial compensation (under Plan B) might be higher than it typically is in years when earnings are high. This outcome might seem perverse to shareholders who do not fully understand the incentive considerations behind the compensation contract. Under plan A, the compensation of managers would appear to be more closely tied to shareholder wealth and would perhaps generate less controversy.

Hedging and the Market's Signal Extraction Problem

As noted earlier, a firm can be hurt if the market undervalues its assets. This occurs, for example, when the firm has a valuable investment opportunity and needs to raise external funds. Corporate hedging has the potential to reduce these "information asymmetries" existing between the firm's managers and the market by improving the "signal-to-noise" ratio in corporate earnings. This can be illustrated by a rather fanciful example. Imagine that a charitable organization hires a fund raiser to solicit donations but is unsure of the fundraiser's ability. Assume that all the organization observes is the amount of money the fundraiser turns in each day. Over time, the organization will gather data to help it resolve the uncertainty concerning the fund raiser's ability. Obviously, a good fundraiser will on average turn in more than a poor one.

Now imagine that each day the fundraiser, before turning in the money, goes to the track and wagers some of the day's proceeds on the horses. The fund raiser then turns in the amount raised plus or minus the winnings or losses at the track. The betting has clearly made it more difficult for the charitable organization to determine the fundraiser's ability. For several days, a good fundraiser could turn in little due to losses at the track while a poor fund raiser might look good due to some lucky bets. In making its assessment of the fundraiser's ability, the organization will put less weight on the daily amounts turned in when these are influenced by the noise of the wagers at the track. A fundraiser who knows he is good and who wants to have this revealed as soon as possible has a clear incentive to avoid the noise added by gambling.

Hedging, to the extent that it removes noise, seems to allow security analysts and others in the market to obtain more precise estimates of the value of a firm's assets. Of course, a key assumption here is that the security analysts do not know all of the risk exposures the firm would face if it did not hedge. If the charitable organization in the example above knows all of the bets placed by the fundraiser at the track and the outcome of each race, then the gambling does not produce noise. Similarly, if the analysts know precisely the foreign exchange exposure, interest rate risk exposure and oil price risk exposure of the company at each moment, then the company gains nothing in terms of eliminating noise by hedging these risks. Of course, if the analysts do not know the company's exposure, the firm's management has the alternative strategy of removing the noise not by hedging but by publicly disclosing the firm's exposure. In a similar manner, the fund raiser need not avoid the track altogether to remove the noise; instead, he could give the charitable organization his track receipts and disclose his betting for the day. The hedging approach might be preferred to the disclosure approach since it puts less of a burden on the market. For many investors, it may be difficult to process all of the information necessary to describe the risk exposures of a large company.[24]

Rewards for Supplying Hedging Services

Our final reason for hedging on corporate account is based on all the above reasons for hedging. These show that corporations can gain by hedging and should in many circumstances be willing to pay another firm or institution to take the counterparty position if necessary. Consider a company, Company A, that is exposed to a particular risk which it has no compelling reason to hedge. This company is in a natural position to provide a hedging contract to another firm, Company B, that is exposed to the same risk but in the opposite way. If Company B derives significant benefit from hedging its exposure, then Company A may be in a position to demand favorable terms of the contract. Whether it can depends on whether there are other potential suppliers of the hedging contract that can compete on the same terms as Company A. The simple point here is that even if the firm has no demand for hedging, its operating exposures may place it in a privileged position to supply hedging services and to receive value for doing so.

Conclusion

We have shown a number of ways in which hedging on corporate account can increase shareholder value. While a firm that hedges on corporate account can change the risk borne by shareholders in their portfolios, the gains from this are likely to be small. The substantial gains produced by hedging are due to the fact that risk affects the expected cashflows corporations can deliver to their shareholders because of taxes, bankruptcy costs, flotation costs for externally generated funds, and other "market imperfections." These considerations make the firm behave as if it were risk averse when it acts in the interest of shareholders and creates a need for hedging. Moreover, for most of these market imperfections, hedging on, shareholders' accounts does not substitute for hedging on corporate account. A shareholder's hedging on his own account cannot lower the firm's expected costs of bankruptcy or financial distress. Nor can a shareholder take a position in a futures market and change the firm's expected tax liability. The firm itself must hedge to capture these advantages of risk reduction. We have not described in any detail how derivatives can be used to hedge since this is done elsewhere. Instead of looking at how derivatives can be used by corporations, we have asked the prior questions of whether they should be used at all and why. The justifications given above for hedging on corporate account show that corporations have a legitimate demand for instruments such as derivatives that they can use to control risk.

References

Brealey, R. A and S. Myers. Principles of Corporate Finance. New York, NY: McGraw-Hili, 1991.

Breeden, D. and S. Viswanathan. "Why Do Firms Hedge? An Asymmetric Information Model," working paper, Duke University, 1990.

DeMarzo, P. and D. Duffie. "Corporate Incentives For Hedging and Hedge Accounting," Working Paper, Northwestern University, 1992.

Froot, K. A, D. S. Scharfstein, and Jeremy c. Stein. "Risk Management Coordinating

Corporate Investment and F'mandng Policies." Journal of Finance 48,5 (1993): 1629-1658.

Modigliani, F. and M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment" American Economic Review 48 (1958): 261-297.

Myers, S. and N. Majluf, "Corporate Financing and Investment Decisions When Firms Have Infonnation That Investors Do Not Have." Journal of Financial Economics 13,2 (1984): 187-221.

Ross, S. A, R. W. Westerfield and J. F. Jaffe. Corporate Finance. Homewood, IL: Irwin, 1993. Smith, C. W., and R. M. StuIz. "The Determinants of a Finn's Hedging Policies." Tournai of

Financial and QuantitativeAnalysis 20,4 (1985): 391-405.,

Stulz. R. M. 1984. "Optimal Hedging Policies." Journal of Financial and Quantitative Analysis 19,2 (1984): 127-140.

___' "Managerial Discretion and Optimal Financing Policies," Journal of Financial Economic._ 260990): 3-27.

Van Home, J. C. Financial Management and Policy. Englewood Cliffs, NJ: Prentice Hall, 1992.

Notes

Since writing this chapter David Fite has left Bankers Trust.

- We thank Anat Admati and Howard Mason for helpful comments.

- For example, Ibbotson and Sinquefield report that from 1926 to 1991, the average return on common stocks was 8.5% higher than the return on U. S. Treasury bills. This is an estimate of the risk premium that investors require to hold risky common stocks. The estimated risk premium for small company stocks over this period is even higher: 14.60/0.
- A diversified portfolio is one that is composed of many securities. Moreover, the value of each security in a diversified portfolio is a small portion of the total value of the portfolio. With the growth of large institutional investors such as pension funds and mutual funds, a large portion of corporate liabilities is now held by diversified investors.
- It is possible for some investors to hold some stocks long and others short in such a way that energy cost risk disappears. The average investor, however, must hold a portfolio that is sensitive to energy cost risk.
- See Modigliani and Miller (1958).
- We assume that the diversified shareholders also hold shares in the firm that exports to Canada and that took the other side of the futures contract. If this is true, the diversified shareholders' exposure to foreign exchange risk remains neutral.
- We have not provided a full argument for why large shareholders exist. Corporate hedging lowers the cost of a large shareholder's taking an undiversified position, but it does not remove that cost entirely. Some other explanation must be provided for large shareholders. It is often suggested that there are benefits to holding controlling stakes in a corporation beyond the cash flows paid out to shares and that this compensates the large shareholder for the disadvantages of a large position. Another explanation is that large. shareholders are caught in undiversified positions by historical circumstances and it is too costly-for them to sell their shares and diversify. This is because large shareholders who sell out might be required to pay significant capital gains taxes. It is also possible that the sale of their shares might have a large negative impact on the market price due in part to the market's recognition that they will no longer monitor the firm as closely. Note also that if the diversified shareholders use corporate hedging to lower the cost of a large shareholder's taking a larger position in the firm, there must be some way that the firm can commit to maintaining a hedging policy in the future. An exploration of these issues is beyond the scope of this article.
- As we mentioned above, it is possible to construct portfolios that are unatfected by pervasive risks. This is done by carefully choosing portfolio weights so that those securities having positive exposures to the pervasive risk exactly offset those having negative exposures. For example, one balances the positive exposure of oil companies to oil price risk with the negative exposure of airlines. If all securities have the same exposure (e.g., they are all affected positively), then this is still possible if one is willing to take short positions. The point is that diversification alone is not sufficient to remove pervasive risks. Moreover, although any individual investor can construct a portfolio that removes a pervasive risk, not all investors can do so.
- One might think that from a diversified investor's point of view, it makes no difference whether a particular firm hedges away a pervasive risk or a nonpervasive risk. It would seem that in both cases the effect on the diversified investor's portfolio is roughly proportional to the weight the firm's stock receives in the portfolio, and in a diversified portfolio this weight is small. This intuition is not correct. Assume, for example, that an investor holds 100 stocks and puts 1 % of his wealth in each. (Dividing the investment equally among the 100 firms is generally not the optimal way to diversify. We make this assumption only to simplify the illustration.) Assume that firm j' s return is equal to F + e_i; where i: measures the effect of the pervasive risk on returns (in this case, the pervasive risk affect all of the firms in the same way), and e_i captures the risk that affects only firm i. For simplicity, assume that the variance of each of the e_i's is equal to V and the variance of the pervasive risk is equal to W . Then if no firm hedges any of its risks, the variance of the diversified investor's portfolio is W + V /100. If the first firm hedges its nonpervasive risk component (i.e., e_1) the variance of the diversified investor's portfolio falls to W + 99V / 100. If instead the first firm hedges its exposure to the pervasive risk (F) and this risk is not transferred to any of the other 99 firms in the diversified investor's portfolio, then the variance of the diversified investor's portfolio falls to $(99/100)^2$W + V / 100. If W is roughly equal to V (the variances of the pervasive and the nonpervasive risks are roughly equal), the hedging of the pervasive risk reduces the variance of the diversified investor's portfolio by 199 times

the amount hedging the nonpervasive risk does. As N (the number of stocks in the diversified investors portfolio) grows, so does the difference between the effects of hedging pervasive and nonpervasive risks.

- We implicitly assume that the pilot's employer is not fully hedged against oil price risk and that when oil prices increase, the pilot's compensation falls. This occurs, for example, if the pilot is temporarily laid off due to a p.ecline in air travel. Issues concerning employee compensation and corporate hedging are discussed below.
- The reader may object that we have assumed that the majority of shareholders are pilots and not oil company employees. After all, pilots are the ones who should shy away from companies with negative exposure to oil price risk. Even though this seems perverse, it could occur if there are many more pilots in the economy than oil industry employees or, more to the point, if pilot's have greater wealth to invest.
- We ignore transactions costs and other fees associated with the mutual fund since these are generally small. Moreover, these only invalidate our assertion about the symmetry of gains and losses in a mutual fund if these costs have a nonlinear relation to the fund's returns.
- For more discussion on the effects of a convex tax structure on the value of corporate hedging, see Smith and Stulz (1985).
- Of course, given the progressive nature of personal taxation in many countries, individual taxpayers who have volatile incomes might reduce their expected tax liability by hedging on personal account. This does not in any way reduce the need for the corporation to hedge on corporate account as a way of reducing corporate tax liability.
- Note that we need to distinguish between what is caused directly by bankruptcy or financial distress and what is caused by general market conditions. The proverbial firm that manufactured buggy whips in the 19205 went out of business because the market for its product changed - not because of its financial structure. Even if the buggy whip firm had no debt at all, it still would go out of business. We are concerned here with what happens when a firm has debt; and because of the failure of the firm's creditors to reorganize the firm quickly and efficiently, the firm follows a different (and lower-value) trajectory than it would have followed if it had no debt.
- It might be argued that the shareholders do not completely absorb the costs of bankruptcy and financial distress but instead that they share them with the bondholders. After all, the bondholders have expenses; and even if these are paid out of the firm's assets, this is money they might otherwise have received. This argument is wrong because it focuses only on what occurs at the time of financial distress. At the time the bonds are issued, the price is set to compensate bondholders for their expected losses due to these costs. The shareholders thus receive less from the bond issue, and the shareholders pay the expected costs of the bondholders at the time of issuance.
- Some additional discussion of the ability of hedging to increase debt capacity can be found in Smith and Stulz (1985).
- This statement is a bit too strong. The NPV rule looks at investment in a "static" environment where investing in a project is a take-it-or-leave-it matter. In a more dynamic context, firms may find it optimal to delay initiating a project with a positive net present value since over time more information becomes available. In this sense, investment projects are like call options for which early exercise is not necessarily optimal. Even if we consider investment in a dynamic context, the story we tell below does not change in any major way.
- This example is based on Myers and Majhir (1984).
- For a more detailed discussion of this rationale for hedging, see Froot, Scharfstein and Stein (1993).
- We implicitly assume here that the risk of success or failure is diversifiable risk, so it is appropriate to consider only the expected outcome. In other words, we assume no need to adjust for risk.
- Some additional discussion on the ability of hedging to reduce agency problems can be found in Stutz (1990).
- See Stu1z (1984) for a discussion of the differences between the diversified position of shareholders and the undiversified position of managers and other employees. Stu1z argues that this creates an incentive for the managers to hedge on corporate account if they are free to do so.
- H all investors In the economy are equally averse to risk, then for optimal risk sharing, each employee of a company should bear a fraction of the company's risk that is equal to the fraction of his wealth to the total wealth in the economy. To illustrate in a f1D'Y rough way the magnitude of this amount, we consider an employee of a company who has $100,000 to Invest. Since the total value of the U. S. stock market is approximately 54 trillion, the investor should bear something on the order of 1/4,000,000 of the company's risk. (This fraction actually overestimates the exposure the employee should face if risk is shared completely since we ignore international diversification and investment in bonds. Since the total value of the world capital market is estimated to be well over 520 trillion, complete risk sharing would put the fraction closer to 1/20,000,000. However, if we also account for noninvestable wealth such as human capital, the employee's wealth increases as does world wealth. Since we only want to establish a rough order of magnitude here, we do not consider these other factors and take the lower value of 1/4,000,000 to obtain a conservative estimate.) Now assume that the company is a $500-million company and suppose that it loses 10% of its value. If risk is efficiently shared, the employee should suffer a loss of only $12.50(= $50,000,000/4,000,000). This means that a 10% loss for the company is only a 0.0125% loss for the employee. It might be argued that it makes no difference how much risk the employee faces in his compensation since the employee can always hedge on his own account to remove this risk. Here

we must again distinguish between pervasive risks and company-specific risks. For the former, the employee can potentially make adjustments in his portfolio to balance his exposure. For example, the Exxon employee who is exposed to oil price risk through his compensation can adjust by holding very little Investment In stocks that have positive exposure to oil prices and by increasing his holdings in those that have negative exposure. In some cases, the employee can also use derivatives and other hedging instruments to control these risks, but for many employees this is costly. If employees are forced to do this, then efficient risk sharing is in all likelihood not being achieved In the least costly manner. While the employee has some ability to manage his exposure to pervasive risks, he has much less ability to control his exposure to company-specific risks. The only effective way for the employee to remove a significant exposure to these risks is through shortselling his employer's stock. When this is allowed. it is generally quite costly for the employee. Of course, the shareholders of a company clearly have legitimate concerns about employees' taking short-sale positions in the stock. especially if these employees are upper-level managers. This means that employees will face restrictions on shortselling. When these restrictions are enforced. employees cannot hedge exposures to company-specific risk in their compensation.

- For models of the use of hedging in improving the outside market's signal extraction problem, see Breeden and Viswanathan (1990) and Demarzo and Duffie (1992).

Notes -

ENGINEERING OF MARKET VOLATILITY

Objectives

- You will be able to show how you can isolate the volatility of a risk factor from other related risks and then constructinstruments that can be used to trade in.

Liquid instruments that involve pure volatility trades are potentially very useful for market participants who have natural exposure to various volatilities in their balance sheet or trading book. The classical option strategies have serious drawbacks in this respect. When trader takes a position or hedges a risk, he or she expects that the random movements of the underlying would have a known effect on the position. The underlying may be random, but the payoff function of a well-defined contract or of a position has to be known. Payoff functions of most classical volatility strategies are not invariant to underlying risks, and most volatility instruments turn out to be imperfect tools for isolating the risk. Even when traders anticipations come true, the trader may realize that the underlying payoff function has changes due to movements in other variables. Hence, classical volatility strategies cannot provide satisfactory hedges for volatility exposures. The reason for this, and possible solutions are the topics.

Until a few years ago, traditional volatility trades involved buying and selling portfolios of call and put options, straddles or strangles, and then possibly delta-hedging these positions. But, such volatility positions were not pure and this led to a search for volatility tools whose payoff function would depend on the volatility risk only. Variance and volatility swaps are tow pure volatility instruments that were developed. This examines these new instruments. They are interesting for at least two reasons. First, volatility is an important risk for market practitioners, and ways of hedging and pricing such risks have to be understood. Second, the discussion of volatility swaps constitutes a good example of the basic principles that need to be followed when devising new instruments.

The uses variance swaps instead of volatility swaps to conduct the discussion. Although, markets, in general, use the term volatility, it is more appropriate for our purpose to perform the calculations with respect to variance, the square of volatility. Variance is the second centered moment of a random variable, and it falls more naturally from the formulas used. For example, volatility (i.e. standard deviation) instruments require convexity adjustments, whereas variance instruments in general do not. Thus, when we talk about vega, for example, we refer to variance vega. This is the sensitivity of the option's price with respect to o^2, not o. In fact, in the heuristic discussion, the term volatility and variance are used interchangeably.

1. Volatility Positions

Volatility positions can be taken with the purpose of hedging a volatility exposure or speculating on the future behavior of volatility. These positions require instruments that isolate volatility risk as best as possible. To motivate the upcoming discussion, we introduce two examples that illustrate traditional volatility positions.

i. Trading Volatility Term Structure

We have seen several examples for strategies associated with shifts in the interest rate term structure. They were called curve steeping or curve flattening trades. It is clear that similar positions can be taken with respect to volatility term structures as well. Volatilities traded in markets come with different maturities. As with the interest rate terms structure, we can buy one "maturity" and sell another "maturity" as the following examples shows.

Example

A. dealer said he was considering selling short-dated 25-delta euro puts/dollar calls and buying a longer-dated straddle. A three-month straddle financed by the sale of two 25-delta one-month puts would have cost 3.9% in premium yesterday.

These volatility plays are attractive because the short-dated volatility is sold for more than the cost of the longer-maturity options.

In this particular example, the anticipations of traders concern not the level of an asset price or return, but instead, the volatility associated with the price. Volatility over one interval is bought using the funds generated by selling the volatility over a different interval.

Apparently, the traders think that short-dated euro volatility will fall relative to the long-dated euro volatility. The question is, to what extent the positions taken will meet the traders' needs, event when their anticipations are borne out. We will see that the payoff function of this position is not invariant to change in the underlying euro/dollar exchange rate.

ii. Trading Volatility across Instruments

Our second example is from the interest rate sector and involves another "arbitrage" position on volatility. The trader buys the volatility of one risk and sells a related volatility on a different risk. This time, the volatilities in question do not belong to different time periods, but instead are generated by different instruments.

Example

Dealers are looking at the spreads between euro cap-floor straddle and swaption straddle volatility to take advantage of a 5% volatility difference in the 7-year area. Proprietary traders are selling a two-year cap-floor straddle starting in six years with vols close to 15%. The trade offers a good pick-up over the five-year swaption straddle with volatility 10%. This compares with a historical spread of closer to 2%.

Cap-floors and swaptions are instruments on interest rates. There are some similarities between them but also differences. Selling a cap-floor straddle will basically be short interest rate volatility, In the example, the traders were able to take this position at 15% volatility. On the other hand, buying a swaption amounts to a long position on volatility. This was done at 10%, which gives a volatility spread of about 5%. The example, states that the latter number has historically been around 2%. Hence, by selling the spread the traders would benefit from a future narrowing of difference between the volatilities to the two instruments.

This position's payoff is not invariant to interest rate trajectories. Even when volatilities behave as anticipated, the path followed by the level of interest rates may result in unexpected payoffs. The following discussion intends to clarify why such positions on volatility have serious weaknesses and require meticulous risk management. We will consider pure volatility positions later.

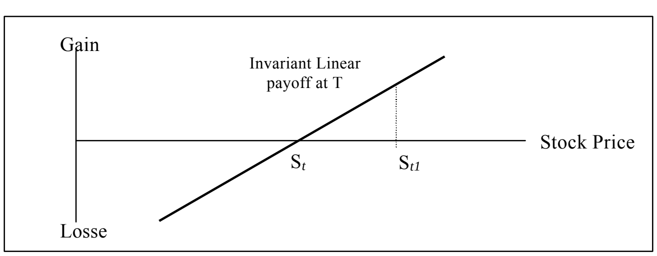
2. Invariance of Volatility Payoffs

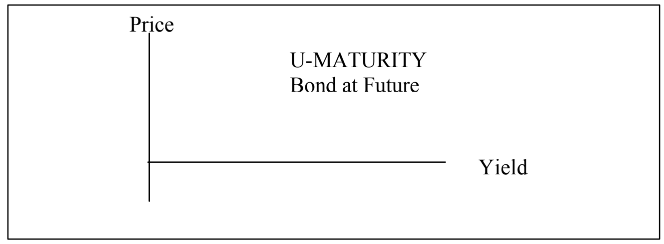
Convexity was used to isolate volatility as a risk. We showed how to convert the volatility of an underlying into "cash," and with that took the first steps toward volatility engineering.

Using the method, a trader can hedge and risk-manage exposures with respect to volatility movements. Yet, these are positions influenced by variables other than volatility. Consider a speculative position taken by an investor.

Let S_t be a risk factor and suppose an investor buys S_t volatility at time t_0 for a future period denoted by $[t_1, T]$, T being the expiration of the contract. As in every long position, this means that the investor is anticipating an increase in realized volatility during this period. If realized volatility during $[t_1, T]$ exceeds the volatility "purchased" at time t_0, the investor will benefit. Thus far this is not very different from other long positions. For example, a trader buys a stock and benefits if the stock price goes up. He or she can buy a fixed receiver swap and benefit if the swap value goes up (swap rate goes down). Similarly, in our present case, we receive a "fixed" volatility and benefit if the actual volatility exceeds this level.

By buying call or put options, straddles, or strangles, and than delts-hedging these positions the trader will, in general, end up with a long position that benefits if the realized volatility increases. Yet, there is one major difference between such volatility positions and positions taken on other instruments such as stocks, swaps, forward rate agreements (FRAs), and so on. Consider, that shows a long position on a stock funded by a money market loan. As the stock price increased, the position benefits by the amount $S_{t1} - S_{t0}$. This potential payoff is known and depends only on the level of S_{t1}. In fact, it depends on S_t linearly. We have a short-dated discount bond position. As the yield decreases, the position gains. Again, we know how much the position will be making or losing, depending on the movements in the yield. $y_{t,}$ if convexity gains are negligible.





A volatility position taken via, say, staddles, is fundamentally diffenent from this as the payoff diagram will move depending on the path followed by variables other than volatility. For example, a change in (1) interest rates, (2) the underlying asset price, or (3) level of implied volatility may lead to different payoffs at the same realized volatility level.

Variance (volatility) swaps, on the other hand, are pure volatility positions. Potential gains or losses in positions taken with these instruments depend only on what happens to realized volatility until expiration. How volatility engineering can by used to set up such contracts and to study their pricing and hedging. We begin with imperfect volatility positions.

3. Imperfect Volatility Positions

In financial markets, a volatility position is often interpreted to be a static position taken by buying and selling straddles, or a dynamically maintained position that uses straddles or options. As mentioned previously, these volatility positions are not the right way to price, hedge, or risk-manage volatility exposure. In this section, we go into the reasons for this. We consider a simple position that consists of a dynamically-hedged single-call option.

3.1 A Dynamic Volatility Position

Consider a volatility exposure taken through a dynamically maintained position using a plain vanilla call. To simplify the exposition, we impose the assumptions of the Black-Scholes world, where there are no dividends, the interest rate, r, and implied volatility, o, are constant, there are no transaction costs, and the underlying asset follows a geometric process. Then the arbitrage free value of a European call $C(S_t, t)$ will be given by the Black-Scholes formula:

Where S_t is the spot price, and K is the strike. The d_i i = 1,2 are given by For simplicity, and without loss of generality, we let

r = 0

This simplifies some expressions and makes the discussion easier to follow.

Now consider the following simple expeiment. A trader uses the Black-Scholes setting to take a dynamically-hedged long position on implied volatility. Implied volatility goes up. Suppose the trader tracks the gains and losses of the position using the corresponding variance vega. What would be this trader's possible gains is the following specific case? Consider the following simple setup.

1. The parameter of the position are as follows:

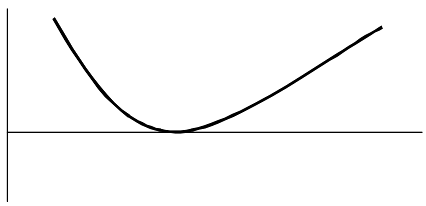
   Time to expiration = .1

   $K = S_{t0} = 100$

$\sigma = 20\%$

Initially we let $t_0 = 0$.

2. The trader expects an increase in the implied volatility from 20% to 30%, and considers taking a long volatility position.
3. To buy into a volatility position, the trader borrows an amount equal to 100 C (S_t,t) , and buys 100 calls at time t_0 with funding cost $r = 0^4$.
4. Next, the position is delta-hedged by short-selling C_s units of the underlying per call to obtain the familiar exposure.



Losses

In this example, there are about 1.2 months to the expiration of this option, the option is at-the-money, and the initial implied volatility is 20%.

It turns out that in this environment, even when the trader's anticipations are borne out, the payoffs from the volatility position may vary significantly, depending on the path following by St. The implied volatility may move from 20% to 30% as anticiapted, but the position may not pay the expected amount. The following example displays the related calculations.

Example

We can calculate the relevant Greeks and payoff curves using Mathematical. First, we obtain the initial price of the call as

$C(100, t_0) = 2.52$

Multiplying by 100, the total position is worth $252. Then, we get the implied delta of this position by first calculating the St-derivative of C(St,t) evaluated at St0 = 100, and then multiplying by 100:

$$100 \left[\frac{aC(S_t, t)}{aSt} \right] = 51.2$$

Hence, the position has +51-delta. To hedge this exposure, the trader needs to short 51 units of the underlying and make the net delta exposure approximately equal to zero.

Next, we obtain the associated gamma and the (variance) vega of the position at t0. Using the given data, we get

$$\text{Variance vega} = 100 \left[\frac{\partial C(S_t, t)}{\partial \sigma^2} \right] = 3152$$

The change in the option value, given a change in the (implied) variance, is given approximately by

$$100\,[\partial C(S_t, t) \cong (3152)\_\partial\sigma^2$$

This means that, everything else being constant, if the implied volatility rises suddenly from 20% to 30%, the instantenous change in the option price will depend on the product of these numbers and is expected to be

$$100\,[\partial C(S_t, t)] = 3152\,(.09 - .04)$$
$$= 157.6$$

In other words, the position is expected to gain about $158, if everything else remained constant.

The points is that the trader was long implied volatility, expecting that it would increase, and it did. So if the volatility does go up from 20% to 30%, is the trader guaranteed to gain the $157.6? Not necessarily. Let us see why not.

Even in this simplified Black-Scholes world, the (variance) vega is a function of St, t, r, as well as o2. Everything else is not constant and the St may follow any conceivable trajectory. But, and this is the important point, when St changes, this in turn will make the vega change as well. The following table shows the possible values for variance vega depending on the value assumed by St, within this setting.

St	Vega
80	0.0558
90	7.4666
100	31.5234
110	10.6215
120	0.5415

Thus, if the expectations of the trader are fulfilled, the implied volatility increases to 30%, but, at the same time, if the underlying price moves away from the strike, say to St1 = 80, the same calculation will become approximately

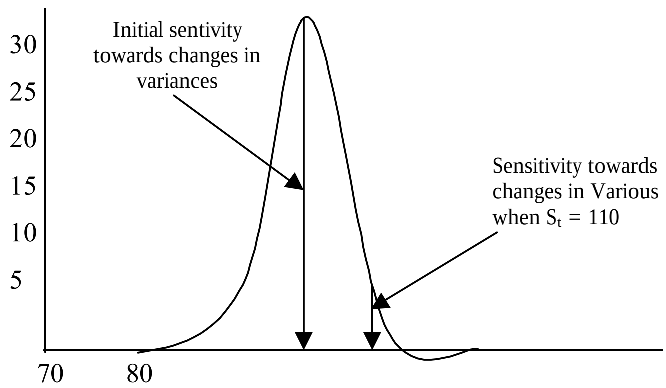
$$\text{Vega}\,(\partial\sigma^2) \cong 5.6(.09 - .04)$$
$$= 0.28$$

Hence, instead of an anticipated gain of $157.6, the trader could realize almost no gain at all. In fact, if there are costs to maintaining the volatility position, the trader may end up losing money. The reason is simple: as St changes, the option's sensitivity to implied volatility, namely the vega, changes as well. It is a function of St. As a result, the outcome is very different from what the trader was originally expecting.

For a more detailed view on how the position's sensitivity moves when St changes, consider we plot the partial derivative:

$$100 \frac{\partial \text{Variance vega}}{\partial S_t}$$

Under the present conditions, we seen that as long as St remains in the vicinity of the strike K, the trader has some exposure to volatility changes. But as soon as St leaves the neighborhood of K, this exposure drops sharpely. The trader may think he or she has a (variance) volatility position, but, in fact, the position costs money, any may not have any variance exposure as the underlying changes right after the trade is put in place. Thus, such classical volatility positions are imperfect ways of putting on volatility trades or hedging volatility exposures.



3.2 Volatility Hedging

The outcome of such volatility positions may also be unsatisfactory if these positions are maintained as a hedge against a constant volatility exposure in another instrument. According to what was discussed, movements in St can make the hedge disappear almost completely and the trader may hold a naked volatility position in the end. An institution that has volatility exposure may use a hedge only to realize that the hedge may be slipping over time due to movements unrelated to volatility fluctuations.

Such slippage may occur for more reasons than just a change in St. In reality, there are also (1) smile effects, (2) interest rate effects, and (3) shifts in correlation parameters in some instruments. Changes in these can also cause the classical volatility payoffs to move away from initially perceived levels.

3.3 A Static Volatility Position

If a dynamic delta-neutral option position loses its exposure to movements in o2 and, hence, ceases to be useful as a hedge against volatility risk, do static positions fare better?

A classic position that has volatility exposure is buying (selling) ATM straddles. Using the same numbers as above, the joint payoff of an ATM call and an ATM put struck at K = 100. This position is made of two plain vanilla options and may suffer from a similar defect. The following example discusses this in more detail.

Example

As in the previous example, we choose the following numerical values:

S_{t0} = 100, r = 0,T - to =.1

The initial volatility is 20%, which means that

$$\sigma^2 = .04$$

We again look at the sensitivity of the position with respect to movements in some variables of interest. We calculate the variance vega of the portfolio

$$V(S_t, t) = 100 \{\text{ATM Put} + \text{ATM Call}\}$$

by taking the partial:

$$\text{Straddle vega} = 100 \frac{\partial V(St,t)}{\partial \sigma^2}$$

Then, we substitute the appropriate values of St, t, o2 in the formula. Doing this for some values of interest for St, we obatian the following sensitivity factors:

St	Vega
80	11
90	1493
100	6304
110	2124
120	108

According to these numbers, if St stays at 100 and the volatility moves from 20% to 30%, the static position's value increases approximately by

$$\partial \text{ Straddle} \cong 6304\,(.09 - .04)$$
$$= 315.2$$

As expected, this return is about twice as big as in the previous example. The straddle has more sensitivity to volatility changes. But, the option's responsiveness to volatility movements is again not constant, and depends on factors that are external to what happens to volatility. The table shows that if St moves to 80, then even when the trader's expectation is justified and volatility moves from 20% to 30%, the position's mark-to-market gains will go down to about 0.56.

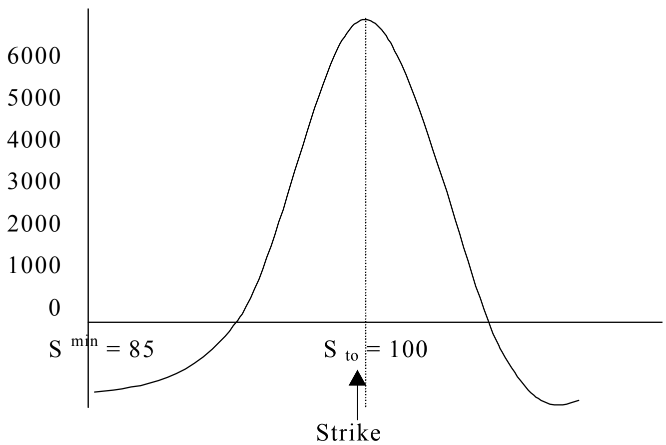
The behavior of the straddle's sensitivity with respect to implied volatility for different values to St. We see that the volatility position is not invariant to changes in external variables. However, there is one major difference from the case of a dynamically maintained portfolio. Static non-delta-hedged positions using straddles will benefit from actual (realized) movements in St. For example, if the St stays at 80 until expiration date T, the put leg of the straddle would pay 20 and the static volatility position would gain. This is regardless of how the vega of the position changed due to movements in St over the inverval [t0,T].

4. Pure Volatility Positions

The key to finding the right way to hedge volatility risk or to take positions in it, is to isolate the "volatility" completely, using existing liquid instruments. In other words, we have to construct a synthetic such that the value of the synthetic changes only when "volatility" changes. This position should no be sensitive to variations in variables other than the underlying volatility. The exposure should be invariant. Then, we can use the synthetic to take volatility exposures or to hedge volatility risk. Such volatility instruments can be quite useful.

First, we know that by using options with different strikes we can essentially create any payoff that we like - if options with a broad range of strikes exist and if markets are complete. Thus, we should, in principle, be able to create pure volatility instruments by using judiciously selected option portfolios.

Second, if an option position's vega drops suddenly once St moves away from the strike, then, by combining options of different strikes appropriately, we may be able to obtain a portfolio of options whose vega is more or less insensitive to movements n St. Heuristically speaking, we can put together small portions of smooth curves to get a desired horizontal line.



When we follows these steps, we can create pure volatility instruments. Consider the plot of the vega of three plain vanilla European call options, two of which are out-of-the-money. The options are identical in all respects, except for their strikes K0 = 100, K1 = 110, K = 120 are plotted. Note that each variance vega is very sensitive to movements in St, discussed earlier. Now, what happens when we consider the portfolio made of the sum of all three calls ? The sensitivity of the portfolio,

$$V(S_t, t) = \{C(S_t,t,K_0) + C(S_t, t,K_1) + C(S_t,t,K_2)\}$$

again varies as St changes, but less. So, the direction taken is correct except that the previous portfolio did not optimally combine the three options. In fact, we should have combined the options by using different weights that depend on their respective strike price. The more out-of-the-money the option is, the higher should be its weight, and the more it should be present in the portfolio.

Hence, consider the new portfolio where the weights are inversely proportional to the square of the strike K,

$$V(S_t, t) = 1/K_0^2\, C(S_t,t,K_0) + 1/K_1^2\, C(S_t,t,K_1) + 1/K_2^2\, C(S_t,t,K_2)$$

The variance vega of this portfolio that uses the parameter values given earlier. Here we consider a suitable 0 < e, and the range

$$K_0 - e < S_t < K_2 + e$$

The vega of the portfolio is approximately constant over this range. This suggests that more options with different strikes can be added to the portfolio, weighting them by the corresponding strike prices. In the example below we show these conditions.

Example

Consider the portfolio

$$V(St,t) = \frac{1}{80^2} C(S_t,t,80) + \frac{1}{90^2} C(S_t,t,90) + \frac{1}{100^2} C(S_t,t,100) + \frac{1}{110^2} C(S_t,t,110) + \frac{1}{120^2} C(S_t,t,120)$$

This portfolio has an approximately constant vega for the range

$$80 - e < St < 120 + e$$

By including additional options with different strikes in a similar fashion, we can lengthen this section further.

We have, in fact, found a way to create synthetics for volatility positions using a portfolio of liquid options with varying strikes, where the portfolio options are weighted by their respective strikes.

4.1 Practical Issues

In our attempt to obtain a pure volatility instrument, we have essentially followed the same strategy that we have been using all along. We constructed a synthetic. But this time , instead of matching the cash flows of an instrument, the synthetic had the purpose of matching a particular senstivity factor. It was put together so as to have a constant (variance) vega.

Once constant vega portfolio is found, the payoff of this portfolio can be expressed as an approximately linear function of σ^2

$$V(\sigma^2) = a0 + a1\sigma^2 + \text{small}$$

With

$$A1 = \frac{\partial V(\sigma^2, t)}{\partial \sigma^2}$$

as long as St stays within the range:

$$S^{min} = K_0 < S_t < K_n = S^{max}$$

Under these conditions, the volatility position will look like any other long (or short) position, with a positive slope a1.

The portfolio with a constant (variance) vega can be constructed using vanilla European calls and puts. The rules concerning synthetics discussed earlier apply here also. It is important that elements of the synthetic be liquid calls and puts have to be selected. The previous discussion referred only to calls. Practical applications of the procedure involve puts as well. This brings us to two somewhat complicated issues. The first has to do with the smile effect. The second concerns liquidity.

4.1.1. The Smile Effect

Suppose we form a portfolio at time t_0 that has a constant vega as long as S_t stays in a reasonable range.

$$S^{min} < S_t < S^{max}$$

Under these conditions, the portfolio consists of options with different 'moneyness" properties and the volatility smile. In general, as K decreases the implied s(K) would increases for constant St. Under these conditions, the trader needs to accurately determine the smile and the way to model it before the portfolio is formed.

4.1.2 Liquidity Problems

From the previous it follows that we need to select out of the money options for the synthetic since they are more liquid. But as time passes, the moneyness of these changes and this affects their liquidity. Those options that become in the money are now less liquid. Other options that were not originally included in the synthetic become more liquid. Even though the replicating portfolio was static, the liquidity of the constituent options may become a drawback in case the position needs to be unwound.

5. Volatility Swaps

One instrument that has invariant exposure to fluctuations in (realized) volatility is the volatility swap. In this section, we introduce this concept and in the next, we provide a simple framework for studying it.

A variance is, in many ways, just like any other swap. The parties exchange floating risk against a risk fixed at the contract origination. In this case, what is being swapped is not an interest rate or return on an equity instrument, but the volatility that correspond to various risk factors.

In the following section we move to a more technical discussion of volatility (variance)swaps. However, we emphasize again that the discussion will proceed using the variance, rather than the volatility as the underlying

5.1 A Framework for Volatility Swaps

Let S_t be the underlying price. The time T_2 payoff V (T_1, T_2) of a variance swap with a national amount N, is given by the following:

$$V\,(T_1, T_2) = [\sigma^2 T_1 T_2 - F_2 t_0]\,(T_2 - T_1)N$$

Where $\sigma T_1, T_2$ is the realized volatility rate of S_t during the interval $t\varepsilon[T_1, T_2]$ with $t < T_1 < T_2$. It is similar to a floating rate and will be observed only when time T_2 arrives. The F_{to} is the "fixed" St volatility rate that is quoted at time t_0 by markets. This has to be multiplied by $(T_2 - T_1)$ to get the appropriate volatility for the contract period. N is the notional amount that needs to be determined at contract initiation. At time to, the $V(T_1, T_2)$ is unknown. The swap is set so that time t_0 ""expected value" of the payoff, denoted by $V(t_0, T_1, T_2)$ is zero. At initiation, no cash changes hands:

$$V(t_0, T_1, T_2) = 0$$

Thus variance swaps are similar to a vanilla swap in that a "floating" $\sigma^2_{T1,\,T2}\,(T_2 - T_1)$ N is received against a "fixed" $(T_2 - T_1)F^2_{to}$N.

The cash flows implied by a variance swap are shown in figure. The contract is initiated at time t_0, and the start date T_1. It matures at T_2. The "floating" volatility) is the total volatility (variance) of S_t during the entire period $[T_1, T_2]$. F_{to} has the subscript t_o, and hence has to be determined at time to. We look at the two legs of the swap in more detail.

5.1.1. Floating Leg

Volatility positions need to be taken with respect to a well defined time interval. After all the volatility rate is like an interest rate. It is defined for specific time interval. Thus, we subdivide the period $[T_1, T_2]$ into equal subintervals, says, days:

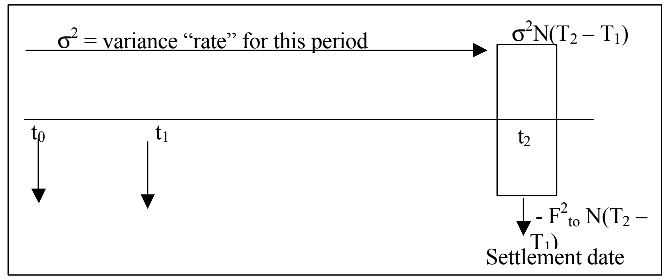
$$T_1 = t_1 < t_2 \ldots\ldots\ldots t_n = T_2$$

with

$$t_2 - t_{i-1} = \delta$$

and then define the realized variance for period δ as

$$\sigma^2_{t1}\delta = \left(\frac{S_{t1} - S_{ti-1}}{S_{ti-1}} - \mu\delta\right)^2$$



Where i = 1 ……n[6]. Hence μ is the expected rate of change St during a year. This parameter can be set equal to zero or any other estimated mean. Regardless of the value chosen. μ is zero, then the right hand side is simply the squared returns during intervals of length δ.

Adding the marginal variances for successive intervals $\sigma^2_{T1,\,T2}$ is equal to

$$(\sigma^2_{T1,\,T2})\ T_2 - T_1) = \Sigma \left(\frac{St1 - Sti\text{-}1}{S_{t1}} - \mu\delta\right)^2$$

Thus σ2T1T2 represents the realized percentage variance of the St during the intervals [T1, T2].

If the intervals become smaller and smaller, $\delta \rightarrow 0$, the last expression can be written as,

This formula defines the realized volatility (variance). It is a random variable at time t_0 and can be viewed as the floating leg of the swap. Obviously, such floating volatilities can be defined for any interval in the future and can then be exchanged against a "fixed" leg.

5.1.2. Determining the Fixed Volatility

Determining the fixed volatility, F_{t0} , will give the fair value of the variance swap at time $t_{0.}$ How do we obtain the numerical value of F_{t0} ? We start by noting that the variance swap is designed so that its fair value at time is equal to zero. Accordingly, the $F^2{}_{t0}$ is that number (variance) , which makes the fair value of the swap equal zero. This is a basic principle used throughout the text and it applies here as well.

We use the fundamental theorem of asset pricing and try to fund a proper arbitrage-free measure P such that What could this measure P be? Suppose markets are complete.

We assume that the continuously compounded risk-free spot rate r is constant. The random process $\sigma^2_{T1..T2}$ is, then, a nonlinear function of S_u T_1

$< u < T_2$ only:

Under some conditions, we can use the normalization by the money market account and let P be the risk-neutral measure.

Then, from Equation taking the expectation inside the brackets and arranging, we get

Therefore, to determine F^2_{t0} we need to evaluate the expectation under the measure P of the integral of σ^2_t. The discrete time equivalent of this is given by

Given a proper arbitrage-free measure, it is not difficult to evaluate this expression. One can use Monte Carlo or tree methods to do this once the arbitrage-free dynamics is specified.

5.2 A Replicating Portfolio

The representation using the risk-neutral measure can be used for pricing. But, how would we hedge a variance swap? To create the right hedge, we need to find a replicating portfolio. We discuss this issue an alternative setup. This alternative has the side advantage that, financial engineering interpretation of some mathematical tools is clearly displayed. The following model starts with Black-Scholes assumptions.

The trick in hedging the variance swap lies in isolating $\sigma^2_{T1,T2}$ in terms of observable (traded) quantities. This can be done by obtaining a proper synthetic. Assume a diffusion process for S_t:

Where W_t is a Wiener process defined under the probability P. The diffusion parameter $\sigma(S_t,t)$ is called local volatility. Now consider the nonlinear transformation.

We apply Ito's lemma to set up the dynamics (i.e. the SDE) for this new process Z_t:

which gives

where the S^2_t term cancels out on the right-hand side. Collecting terms, we obtain

Notice an interesting result. The dynamics for dS_t/S_t and d S_t log are almost the same except for the factor involving $\sigma(S_t,t)^2dt$. This means that we can subtract the two equations from each other and obtain.

This operation has isolated the instantaneous percentage local volatility on the right-hand side. But, what we need for the variance swap is the integral of this term. Integrating both sides we get

We now take the integral on the left-hand side

we use this and rearrange to obtain the result:

We have succeeded in isolating the percentage total variance for the period $[T_1, T_2]$ on the right-hand side. Given that S_t is an asset that trades, the expression on the left-hand side replicates this variance.

5.2 The Hedge

The interpretation of the left hand side in the Equation is quite interesting. It will ultimately provide a hedge for the variance swap. In fact, the integral in the expression is a good example of what Ito integrals often mean in modern finance. Consider

How do we interpret this expression?

Suppose we would like to maintain a long position that is made of $^1/S_t$ units of S_t held during each infinitesimally short interval of size dt, and for all t. In other words, we purchase $^1/S_t$ units of the underlying at time t and hold them during and infinitesimal interval dt. Given that at time t, S_t is observed, this position can easily be taken. For example, if $S_t = 100$, we can buy 0.01 units of S_t at a total cost of 1 dollar. Then, as time passes, S_t will change by dS_t and the position will gain or lose dS_t dollars for every unit purchased. We readjust the portfolio since, the S_{t+dt} will presumably be different, and the portfolio needs to be $^1/S_{t+dt}$ units long.

The resulting gains or losses of such portfolios during an infinitesimally small interval dt are given by the expression

Proceeding in a similar fashion for all subsequent intervals dt, over the entire period $[T_1,T_2]$, the gains or losses of such a dynamically maintained portfolio add up to

The integral, therefore, represents the trading gains or losses of a dynamically maintained portfolio.

The second integral on the left-hand side of Equation is taken with respect to time t, and is standard integral. It can be interpreted as a static position. In this case, the integral is the payoff of a contract written at time T_1, which pays, at time T_2, the difference between the unknown log (S_{T2}) and the known log (S_{T1}). This is known as a log contract. The long and short positions in this contract are logarithmic functions of S_t.

In a sense, the left-hand side of Equation provides a hedge of a variance contract. If the trader is short the variance swap, he or she would also maintain a dynamically adjusted long position on S_t and be short a static log contract. This assumes complete markets.

6. Some Uses of the Contract

The variance (volatility) swaps are clearly useful for taking positions with volatility exposure and hedging. But, each time a new market is born, there are usually further developments beyond the immediate uses. We briefly mention some further applications of the notions developed.

First of all, the F^2_t, which is fixed leg of the variance swap, can be used as a benchmark in creating new products. Its is important to realize however, that this price was obtained using the risk-neutral measure and that is not necessarily an unbiased forecast of future volatility (variance) for the period $[T_1, T_2]$. Just like the FRA market prices, the F_t will include a risk premium. Still, it is the proper price to write volatility options on.

The pricing of the variance swap does not necessarily give a volatility that will equal the implied volatility for the same period. Implied volatility comes with a smile and this may introduce another wedge between F_t and the ATM volatility.

Finally, the F^2_t should be a good indicator for risk-managing volatility exposures and also options books.

The following reading illustrates the developments of this market.

Example

A striking illustration of the increasing awareness of volatility among the hedge fund community is the birth of pure volatility funds. But just as notable as the introduction of specialist volatility investment vehicles is the growing realization among regular directional hedge funds of the need to manage their volatility positions.

"As people become aware of volatility, they are increasingly looking to hedge or trade the vega," said a participant from a directional hedge fund.

Convertible arbitrage funds have also been getting in on the act as they come to fully understand the concept of vega. Volatility is a major factor in the pricing of convertible bonds.

Investments banks have responded to and increased hedge fund interest in volatility by providing new straightforward volatility structures.

The best example of the new breed of simple volatility products is the volatility swap. These are cast-settled forward bets on market volatility which allow the investor to set up a pure volatility trade with a dealer. When the customer sells volatility, the dealer agrees to pay a fixed volatility rate on a national amount for a certain period. In return, the investor agrees to pay the annualized realized volatility for the S&P500 for the life of the swap.

At maturity, the two income streams are netted and the counterparties exchange the difference in whichever direction is appropriate. This type of product encourages hedge fund volatility activity because it offers them a simpler method of trading vega.

Normal volatility trades, such as caps and floors, leave investors exposed to underlying price risk. As the market moves towards the strike price, the gamma effect in hedging the position may cause the investor to lose more on the hedging than the makes on the volatility rate. Careful book management is necessary to control this risk. Most directional hedge funds have so many things to look at that they haven't always got the time, inclination or understanding to trade volatility using the traditional products. "Volatility swaps turn vega into something that people can easily grasp and manage," said one directional hedge fund commentator (IFR, December 31, 1998)

Volatility trading, volatility hedging, and arbitraging all fall within a sector that is still in the process of development.

6. Which Volatility ?

In this dealt with four notions of volatility. These must be summarized and distinguished clearly before we move on the discussion of the volatility smile.

When market professionals use the term "volatility", chances are they refer to Black-Schole's implied volatility. Otherwise, they will use terms such as realized or historical volatility. Local volatility and variance swap volatility are also part of the jargon. Finally, cap-floor volatility and swaption volatility are standard terms in financial markets.

Implied volatility is simply the value of s that one would plug into the Black-Scholes formula to obtain the fair market value of a plain vanilla option as observed in the markets. For this reason, it is more correct to call I t Black-Scholes implied vol or Black volatility in the case of interest rate derivatives. It is quite conceivable for a professional to use a different fomula to price options, and the volatility implied by this formula would naturally be different. The term implied volatility, is thus, a formula-dependent variable.

We can attach the following definitions to the term "volatiltiy".

* First, there is the class of realized volatilites. This is closest to what is contained in statistics courses. In this case, there is an observed or to-be-observed data set, a "sample", $\{x_1 \ldots . x_n\}$, which can be regarded as realization of a possibly vector-stochastic process, x_t, defined under some real-world probability P. This process x_t has a second moment

We can devise an estimator to estimate this σ_t. For example, we can let

Where x_t^m is the m-period sample mean:

Such volatilities measure the actual real –world fluctuations in asset prices or risk factors. One example of the use of this volatility. The σ^2_t defined earlier represented the floating leg of the variance swap discussed here.

- The next class is implied volatility. There is an observed market price. The market practitioner has a pricing formula (e.g., Black-Scholes) or procedure (e.g., implied trees) for this price. Them, implied volatility is that "volatility" number, or series of numbers, which must be plugged in the formula in order to recover the fair market price. Thus, let $F(S_t, t, r, \sigma_t, T)$ be the Black-Scholes price for a European option written on the underlying S_t, with interest rates r and expiration T. At time t, σ_t represents the implied volatility if we solve the following equation (nonlinearly) for σ_t:

  $F(S_t, t, r, \sigma_t, T)$ = Observed price

  This implied volatility may differ from the realized volatility significantly, since it incorporates any adjustments that the trader feels he or she should make to expected realized volatility. Implied volatility may be systematically different than realized volatility if volatility is stochastic and if a risk premium needs to be added to volatility quotes. Violations of Black-Scholes assumptions may also cause such a divergence.

- Local volatility is used to represent the function $\sigma(.)$ in a stochastic differential equation:

  However, local volatility has a more specific meaning. Suppose options on S_t trade in all strikes, K, and expirations T, and that the associated arbitrage free prices, $\{C(S_t, t, K, T)\}$ are observed for all K, T. Then the function $\sigma(S_t,t)$ is the local volatility, is the local volatility if the corresponding SDE successfully replicates all these observed prices either through a Monte Carlo or PDE pricing method.

  In other words, local volatility is a concept associated with calibration exercises. It can be regarded as a generalization of Black-Scholes implied volatility. The implied volatility replicates a single observed price through Black-Scholes formula. The local volatility, on the other hand, replicated an entire surface of options indexed by K and T, through a pricing method. As a result, we get a volatility surface indexed by K and T, instead of a single number as in the case of Black –Scholes implied volatility.

- Finally, we encountered the variance swap volatility. This refered to the expectations of the average future squared deviations. But, because the expectations use the risk-neutral measure, it is different from real-world volatility.

Discussions of the volatility smile relate to these volatility notions. The implied volatility is obviously of interest to most trader but it cannot exist independent of realized volatility. It is natural to expect a close relationship between the two concepts. Also, as volatility trading develops, more and more instruments are written that use the realized volatility as some kind of underlying risk factor for creating new products. The variance swap was only one example.

Notes -

HOW DO CREDIT DERIVATIVES CHANGE FINANCIAL ENGINEERING

Objectives

- You can understand in this lesson a different type of credit derivatives and how it can play a great role in financial engineering.

Dear Friends!

Credit derivatives have had a revolutionary effect on financial engineering. Liquid credit derivatives are the last pieces in the puzzle that a market practitioner needs in order to create practical synthetics for almost any instruments. Without credit derivatives, creating exact synthetics for non-AAA-rated instruments would be possible, but it would be imperfect. A synthetic that does not use credit derivatives would require some effort in modeling credit spreads and would be ad hoc to some extent. The principle that is applied throughout this text is that pricing, hedgeing and risk management should be based on liquid and tradable securities prices as much as possible. With credit derivatives, the ad hoc modeling aspects are minimized, and the model parameters can be calibrated to liquid markets.

Liquid credit derivatives markets extend the creation of synthetics to assets with default risk. Pricing credit is left to the markets. This way credit can be traded separately. In contrast, traditional approaches to credit risk use some ad hoc estimates of credit curves. By adding a proper credit spread to say, the Treasury curve, practitioners obtain discount factors that incorporate credit risk. However, in some states of the world where credit spreads can change 200 to 300 basis points in a rather short time, a correct calibration to the Treasury curve may be much less important than obtaining market information on that particular credit per se, directly from the markets. This is true even though credit spread fluctuations during 'normal' times may be due to fluctuations of the swap spread. These questions and the need to price default separately using liquid instruments, are of fundamental importance to a risk manager and credit derivatives will play a role here.

In this chapter, we first briefly summarize the major credit instruments and then discuss how credit derivatives can be integrated nicely in the engineering of these products. The chapter can only discuss the basis aspects of this important sector. We concentrate on credit default swaps (CDS). This is the primary instruments of credit markets, and it is becoming more so as time passes. We will see that CDSs are a natural extension of liquid fixed-incoming instruments. The chapter will also review the major aspects of some other credit derivatives.

2. Terminology and Definitions

First, we need to define some terminology. The credit sector is relatively new in modern finance although an ad hoc treatment of it has existed as long as banking itself. Some of the terms used in this sector come from swap markets, but others are new and specific to the credit sector. The following list is selective.

1. Reference asset. The instrument on which credit risk is traded. Note that the credit sector adopts a somewhat more liberal definition of the basis risk. A trader may be dealing in loans but may hedge the credit risk using a bond issued by the same credit.
2. Credit event. Credit risk is directly or indirectly associated with some specific events (e.g. defaults, or downgrades). These are important, discrete events, compared to market risk where events are relatively small and continuous.[1] The underlying credit event needs to be defined carefully in credit derivatives contracts. For example, the industry is still in the process of developing the exact definition of a default. A downgrade by a rating agency, on the other hand is less ambiguous in the sense. Recently, there has been some debate on whether restructuring the debt constitutes a credit event[2]. Interestingly, the industry is in favor of not considering restructuring as a credit event. Such controversies are helpful for understanding credit derivatives properly. But they also illustrate that the credit sector is in a transition period during which contracts and documents are settling down.
3. Projection buyer. This is the entity that buys a credit instruments such as a CDS. This entity will make periodic payments in return for compensation in the event of default. A protection seller is the entity that sells the CDS.
4. Recovery value. If default occurs, the payoff of the credit instrument will depend on the recovery value of the underlying asset at the moment of default. This value is rarely zero. Some positive amount will be recoverable. Hence, the buyer needs to buy protection over and above the recoverable amount. Major rating agencies such as Moody's or Standard and Poor's have recovery rate tables for various credits. These tables are prepared using past default data.

The credit sector has other sector specific the terms that we will introduce during our discussion. We now briefly review basic credit derivatives. We will concentrate on credit default swaps, as these are by far the most active contracts. However, there is enough interest in other instruments that a brief discussion of these is warranted.

2.1 Types of Credit Derivatives

Crude forms of credit derivatives have existed since the beginning of banking. These were not liquid and, in general, did not possess the desirable of modern financial instruments that facilities their use in financial engineering. Banking services such as a letter of credit banker's acceptances, and guarantees are precursors of modern credit instruments and can be found in the balance sheet of every bank around the world. Traditional credit instruments do not lend themselves to convenient uses in financial engineering. Modern credit instruments have their own special characteristics, and in this chapter, we are essentially concerned with these.

Broadly speaking, there are three major categories of credit derivatives (1) Credit event related products make payments depending on the occurrence of a mutually agreeable event. The credit default swap is the most common type here. (2) Credit spread products are those whose payoffs depends on how a particular credit spread changes. An example would be a credit spread option that makes a payment in case a credit spread beyond a strike level K. (3) Mixtures, the most popular being the total return swap (TRS), whose payoffs depend on the behavior of spreads as well as on events such as defaults.

2.1.1. Credit Default Products

Credit risk can be broadly grouped into two different categories. On the one hand, there is credit deterioration. Changes in the underlying credit spread can indicate how credit deteriorates. The second element of credit risk is default risk. The latter is separate from credit deterioration, although it is certainly correlated with it. Default products trade default risk by separating it from credit deterioration risk.

Default products share the properties of instruments that have existed for a long time. For example, banks have issued letters of credit, guarantees, and instruments. The major distinguishing characteristics of these traditional instruments is that they transfer credit risk only. They do not in particular transfer market risk or the risk of credit deterioration. Essentially a payment is made when default occurs. With these products no compensation changes hands when the underlying credit deteriorates. New credit default products share some of the properties of these old instruments. There are two kinds (1) In the case of credit default swaps, a fee is paid periodically until default occurs. If there is no default, the protection ends at contract expiration with no other cash exchange (2) Credit default options are similar to credit default swaps, but the fee is paid up front. Both of these instruments involve swapping a fixed fee against a contingent payment in the case of default. Some of the important features products are as follows:

1. The payment is dependent on an event rather than an underlying price, similar to insurance products and unlike other derivatives. This modification increases legal and documentation risk. In this respect, the efforts of the International Swaps and Derivatives Association (ISDA) are relevant. A great deal of effort goes into standardizing the CDS contracts.
2. The determination of recovery values. This is a difficult component of pricing and needs to be specified clearly in credit derivatives contracts.
3. The issue of settlement. In the case of physical delivery, this does not present a major problem. The protection seller will be the legal owner of the defaulted instruments and may take necessary legal steps for the recovery. But if the contract is cash-settled, then neither party has legal recourse to the borrower unless the party owns the underlying credit directly. For this reason, the industry prefers physical delivery, and a large majority of default swaps settle this way.

We will address the additional characteristics of default products when we study credit default swaps in more detail. Now, we briefly consider instruments that trade credit deterioration.

2.1.2 Spread Products

These credit instruments are similar to standard derivatives. Suppose the market provides a credit spread. Then, option and other derivatives can be written on it, just like in case of equity or interest rates. The novelty here is that the spread itself is dependent on the probability of default and this is a nonlinear stochastic process. The underlying theoretical models often, cannot be simply based on Wiener process-driven stochastic differential equations (SDE). Rather the modeling may have to incorporate some elements of a jump process, or, may involve other non-linear.

We can envisage at least two spread instruments. The first is the credit spreads option, where the payoff would be the excess of a credit spread over a strike price. Essentially this would be similar to a standard caplet with the credit spread being the underlying risk instead of a Labor rate. A second type of spread instruments is a credit spread swap. One counterparty pays the credit spread of an issuer against receiving the spread of another issuer. As mentioned earlier, the markets prefers trading default products so that pure spread products make up a much smaller proportion of the credit derivatives sector.

2.1.3. Mixed Products

The main characteristic of these products is that they are instruments written on a mixture of credit and market risk. The two risks are bundled together and then sold to clients. The main type of mixed credit instrument is the total return swap (TRS).

There is an important difference between default products and mixed instruments such as total return swaps. Because credit deterioration and default risks are traded separately in the market credit default products will give the trader an asset during at that time's market value but will not be subject to direct market risk. On the other hand, total return swaps incorporate market risk.

In the next section, we look at the most liquid credit derivatives in more detail. We study the financial engineering of the credit default swaps.

3. Credit Default Swaps

The major instruments of the credit sector is the credit default swap. A typical default swap is shown in Figure 16-1 from the point of view of a protection buyer. The CDS buyer in a particular credit denotedd by i. pays a constant rate called the CDS rate. The CDS expires at time T. The CDS spread is denoted by d_{to} and is set at time t_0. A payment of dt_oäN is made at every t_0. If no default occurs until T, the contract expires without any other payments. On the other hand, if the credit t defaults during (t_0,T) the CDS buyer receives N dollars from the seller. Against this receipt of cash, the protection buyer has to deliver physically eligible debt instruments with par value N dollars. These instruments will be from a deliverable basket, and are clearly specified in the contract at time t_0. Obviously, one of these instruments will in general be cheapest to deliver in the case of default and all players may want to deliver that particular bond.

Later in this chapter we will consider additional properties of the default swap market that a financial engineer should be aware of. At this point, we discuss the engineering aspects of this product. This is especially important, because we will show a default swap will fall naturally as the residual from the decomposition of a typical risky bond. In fact, we will take a risky bond and decompose it into components. The key component will be default swap. This natural function played by default swaps partly explains their appeal and their position as the leading credit instrument.

We discuss the creation of a default swap by using a specific example. The example deals with a special case, but illustrate almost all the major aspects of engineering credit risk. Many current practices involving synthetic collateralized debt obligations (CDOs), credit linked notes (CI.Ns), and other popular credit instruments can be traced to the discussion provided next.

Independently, this section can be seen as another example of engineering cash flows. We show how the static replication methods change when default risk is introduced into the picture. Essentially, the same techniques are used. But, the creation of a satisfactory synthetic becomes possible only if we add CDSs to other standard instruments.

Figure

3.1 Creating a CDS

The steps we intend to take can be summarized as follows. We take bond that has default risk and then show how the cash flows of this bond can be decomposed into simpler liquid constituents. Credit defaults swaps result naturally from this decomposition.

Our discussion leads to a new type of contractual equation that will incorporate credit risk. We then use this contractual equation to show how a credit default swap can be created, hedged and priced in theory. The contractual equation also illustrate some of the inherent difficulties of the hedging and pricing process in practice. At the end of the section, we discuss some practical hedging and pricing issues.

3.2 Decomposing a Risky Bond

We keep the example simple in order to illustrate the fundamental issue more clearly. Consider a risky bond purchased at time t_0, subject to default risk. The bond does not contain any implicit call and put options and pays a coupon c annually over three years. The bond is originally sold at par.

We make two simplifying assumption which can be relaxed with little additional effort. These assumptions do not change the essence of the engineering but significantly facilitate the understanding of the credit instrument. First, we assume that, in the case of default, the recovery value equals zero. Second and without much loss of generality. We assume that the default occurs only in period t_0.

Figure 16.2 shows the cash flows implied by this bond. The bond is initially purchased for 100, three coupon payments are made and the principal of 100 is returned if there is no default. On the other bond, if there is default (in period t_3 only), the bond pays nothing. The optionality due to the default possibility is shown with the fork at time t_3. At time t_3, there are two possibilities and the claim is contingent on these.

How do we reverse engineer these cash flows and convert them into liquid and instrument? We answer the question in steps.

First, we need to introduce a useful trick that will facilitate the application of static decomposition methods to defaultable instruments. We do this in Figure 16-3 Remember that our goal is to isolate the underlying default risk using a single instrument. This risk will be greatly simplified if we add subtract the amount 1000 to the cash flows in the case flows. Yet it is useful for isolating the inherent credit default swap. As we will see.

1. First, there are three coupon payments dates t_1, t_2, and t_3. Of course by assumption, the third coupon payments has default risk. But our “trick” of adding and subtracting100 to the time t_3 cash flows permits considering this as a guaranteed payment at time t_3. as shown in Figure 16.3a. According to the although the last coupon is risky. We can still extract three default free coupon payments from the bond cash flows due to the trick used. In fact, to get the default free coupon payment. We simply pick the positive e at time t3 of Figure t3 of Figure 16.3a. Note that this leaves the negative in cash default occurs.
2. The second type of bond cash flows are the initial and final payment of 100 as shown in Figure 16.3 e. Again the trick of adding and subtracting 100 is used to obtain a default free receipt of 100 at time t_3. These two cash flows are then carried to Figure 16.3e. As a result the negative payment of 100 in the default state t3. remains in Figure 16-3a.
3. Finally Figure 16.3d shows all remaining cash flows. These consist of the negative flow e4 100 that occurs at time t3 default state. This is detached and placed in Figure 16-3 d.

The next step to convert the three cash flows diagrams in Figure 16-3b, 16.3c, and 16-3d into recognizable and preferably liquid contracts in markets. Remember that to do this. We need to ad and subtract arbitrary cash flows to those in Figure 16-3b, 16-3c, and 16-3d while ensuring that the following three conditions are met.

- For each cash flow added, we have to subtract the same amount for its prevent value (or its prevent value) at the same t3, from one of the Figure 16-3, 16-3c, 16-3d.
- These new flows should be introduced to make the resulting instruments liquid.
- When added back together, the modified Figure 16-3, 16-3c, 16-3d should give back the original bond cash flows in Figure 16-3a. This way, we should be able to recover the cash flows of the defaultable bond.

We show this in Figure 16-4. The easiest cash flows to convert into a recognizable instrument are those in Figure 16-3b. If we floating Libor-based payments Lti at times t_1, t_2 and t_3, thee cash flows will look a fixed interest rate swap. This is good because swaps are very instruments. However, one additional modification is required. The fixed-receiver swap rate involves a rate st10, this less than the coupon of a bar bond issued at time t_0. Thus, we have

Figure

$S_{t0} \leq c$

The difference, denoted by d_{to},

$d_{to} = c - s_{to}$

is the credit spread over the swap rate. This is how much a credit, rated A or lower, has to pay over and above the swap rate due to the default possibility. Note that we are defining the credit spread over the corresponding swap rate not over that of the Treasuries. This is, in fact, the market the market practice and the right way to go about calculating the credit spread. It also flows falls naturally from the cash decompositions.

Thus, in order for the cash flows in Figure 16-4a to be equivalent to a receive, we need to subtract d_{to} from each coupon receipt, c, as done in Figure 16-4a. This will make, market the mixed receipts equal the swap rate.

$C - d_{to} = s_{to}$

The resulting cash flows become a true interest rate swap.

This construction leaves an important question unanswered. Where do we place the counterparts of the cash flows d_{to} and Lti, that we just introduced in Figure 16-4a? After all, unless the same cash flows are placed somewhere else with opposite signs, that will cancel out, and the resulting synthetic will not reduce to a risky bond.

A natural place to place the Libor-based cash flows is shown 16-5, Nicely, the addition of the Libor receipts converts the cash flows into a default-free money market deposit with tenor. This deposit will be rolled over at the going floating Libore rate. Note that this is also a liquid instrument.[3]

The final adjustment is how to compensate the reduction of c's by the credit spread dto. Since the first instruments are complete, there is only one place to put the compensating d_{to}'s. We add the dto to the cash flows shown in Figure 16-3 d, and the result in Figure 16-4 b. This is the critical step since, we now obtained a new instrument that has fallen naturally from the decomposition on the risky bond. Essentially this instrument has three receipts of dto, dollars at times t_1 t_2, and t_3. But if default occurs, the instrument will make a payment of e +100 dollars[4].

To make sure that the decomposition is correct. We add Figure 16-4 b and 16-5 vertically and see if the original cash flows are recovered. The vertical sum of cash flows in Figure 16-4a, 16-4n and 16-5 indeed replicates exactly the cash flows of the default bond.

The instrument we have in Figure 16-4 b is equivalent to selling protection against the default risk of the bond. The contract involves collecting fees equal to dto and each t1 until the default occurs. Then the protection buyer is compensated by the loss e100 d_{to}. On the other hand, if there is no default, the fees are collect until the expiration of the contract and no payment is made. We call the instrument a credit swap (CDS).

3.3. A Synthetic

The preceding discussion shows that a defaultable bond can be decomposed into a portfolio made up on (1) a fixed receiver rate swap (2) a default free money market deposit, and (3) a credit default swap. Here, the maturities of the bond. Swap, and the CDs would be the same. The use of the these instruments implies the following equation:

[Defaultable bond on the credit] = [Receiver swap] + [Default – free deposit] + [CDS on the credit]

By manipulating the elements of this equation using the standard rules of algebra, we can obtain synthetics for every instrument in the equation. In the next section we show two applications.

3.4 Using the Contractual Equation

As a first application, we show how to obtain a hedge for a CDS position by manipulating the contractual equation. Second, we discuss the implied pricing and the resulting real world difficulties.

There are, of course many, uses of the preceding contractual equation. For example, using a CDS, we can construct a synthetic syndicated loan, or a corporate bond for any credit that normally does not issue such securities. Some of these will be discussed later in the chapter when we look at the way CDSs are used in the industry.

3.4.1. Creating a Synthetic CDS

First, we consider the way a CDS would be hedged. Suppose a market sells a CDS on a certain name. How would the market hedge this position while it is still on his or her books?

To obtain a hedge for the CDS, all we need to do is to manipulate the contractual equation obtained above. Rearrange we-obtain.

[Defaultable bond issued by the credit] - [Receiver swap] - [Default-free deposit] = [CDS on the credit]

Remembering that a negative sign implies the opposite position in the relevant instruments, we can write the formal synthetic for the credit default swap as.

[CDS on the credit] = [Risky bond on credit] + [Payer swap] + [Default free loan]

The market maker who sold such a CDS and provided protection needs to take the opposite position on the right hand side of this equation. This is to say, the credit derivatives dealer will first short the risky bond, deposit the received 100 in a default-free deposit account, and contract a receiver swap. This and the long CDS position will then "cancel" out. The market maker will make money on the bid-ask spread.

3.4.2 Pricing and Hedging

The second application of the contractual equation, at least theoretically. The contractual equation that leads to the creation of a credit swap can also be used to price the CDS. According to this, in order to obtain the fee for writing protection on this credit, we need to calculate the difference between the yield of the corresponding liquid defaultable bond and the current swap rate.

3.5 Real-World Complications

The contractual equation obtained in this chapter provides a natural hedge for the CDS and shows one way of pricing it. Similar contractual equations may provide usable hedges and pricing methods for some bread-and-butter instruments with negligible credit risk, but for CDSs these equations are, essentially theoretical. The native approach discussed above may sometimes misprice the CDS. The hedge obtained earlier may not hold. There may be several reasons why the benchmark spread[1] on this credit may deviate significantly from the CDS rates. We briefly discuss some of these reasons.

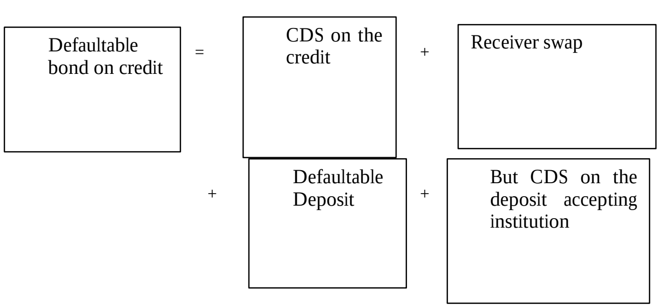
1. In the preceding example, the CDS had a maturity of 3 years. What if the particular credit had no outstanding 3 year bonds at the time CDS was issued? Then, the pricing would be more complicated and the benchmark spread could very well deviate loan the CDS rate.
2. Even if similar maturity bonds exist, these may not be very liquid, especially during times of high market volatility. Then, it would be natural to see discrepancies between the CDS rates and the benchmark spreads.
3. The tax treatment corporate bonds and CDSs are different, and this introduces a wedge between the corresponding spread and the CDS rate.
4. As mentioned earlier, CDSs result in physical delivery in the case of default. But this delivery is from a basket of deliverable bonds. This means that the CDS contains a delivery equation, which was not built into the contractual equation presented earlier.

In reality, another important issue arises. The construction of the synthetic shown above used a money market account was assumed to be risk-free. In general, such money market accounts are almost never risk-free and the deposit accepting institution will have a default risk. This introduces another wedge between the theoretical construction and actual pricing. However, if such is the case, the yield from the money market account is not the default-free Libor L_{t0}, but instead.

$$L_{t0} - \mu_{t0}$$

where μ_{t0} is the CDS rate for the deposit-accepting institution.

Thus additional credit risk that creeps into the construction can, in principle, he eliminated by buying a new CDS for the deposit accepting institution. The cost of this CDS will, by definition be $ì_{t0}$ and adding this CDS to the original contractual equation will solve the problem. The The new contractual equation becomes.



This would give the "correct" synthetic for the bond. Looking at these real-life modifications, it is obvious that actual pricing of CDSs needs to take into account several difficulties that do not exist in the simple engineering example discussed earlier. Yet, the analysis does capture the main point that we want to make in this chapter: namely that the introduction of credit derivatives greatly simplifies the creation of synthetics for almost any standard instrument.

3.5.1 Real-World Difficulties in Hedging

We consider an interesting observation from financial markets. Let b_{t0} be the yield of a defaultable coupon bond that was issued at time t_0 at par. Let d_{t0} be the T-maturity CDS spread on the same name. Finally, let the s_{t0} and T_{t0} be the par swap rate and the treasury yield with maturity T, respectively. Then, the simple contractual equation derived earlier suggests that we should have

$$d_{t0} = b_{to} - S_{t0}$$

After all, if we have the inequality

$$d_{t0} > b_{to} - S_{t0}$$

then, instead of buying credit protection on the issuer, the client would simply short the bond and get a receiver swap. This will provide the same protection against default, and at the same time, cost less.

However, an observation of trading in CDS markets would show that real world data sometimes have the following characteristics.[1]

$$d_{t0} \neq b_{to} - S_{t0}$$

Does this mean that there is an arbitrage positively? In fact, such inequalities can be caused by many different factors. We briefly list these below.

1. CDS protection is "easy" to buy. On the other hand, it is "costly" to short bonds. One has to first go to the repo market to find such bonds, and repo has the mark-to-market properly. With CDS protection, there will be no such inconveniences.
2. Shorting a bond is risky because of the possibility of a short-squeeze. If too many players are short the bond, the position may have to be covered at a much higher price.
3. Some bonds may be very hard to find when a sudden need for protection arises.
4. Also as discussed earlier, a delivery option premium is included in the CDS rate.

These factors may cause the theoretical hedge to be different from the CDS sold to clients. Finally, it should be noted that, when the probability of default becomes significant, CDS dealers may suddenly move their prices out and stop trading.

4. Total Return Swaps

Total Return Swaps (TRS) trade default, credit deterioration, and market risk simultaneously. It is instructive to compare them with CDSs. In the case of a CDS, a protection buyer owns a bond issued by a credit and would like to buy insurance against default. This is done by making constant periodic payments during the maturity of the contract to the protection seller. It is similar to, say fire insurance. A constant amount is paid, and of during the life of the contract the bond issuer defaults, the protection seller compensates the protection buyer for the loss and the contract ends. The compensation is done by paying the protection buyer the face value of, say 100 and then, in return accepting the delivery of a deliverable bond issued by the credit. In brief, CDSs are instruments for trading defaults only.

A total Return Swap has a different structure. Consider a bond or any arbitrary risky security issued by a credit. This security makes two types of payments. First, it pays a coupon interest. Second, there will be associated capital gains (appreciation in asset price) and capital losses (depreciation in asset price), which includes default in the extreme case. In a TRS the protection seller pays any depreciation in the asset price during periodic intervals to the protection buyer. Default is included in these payments, but it is not the only component. In general, assets gain or lose value for many reasons, and this does not mean the issuer has defaulted or will default. Nevertheless, the protection buyer for those losses as well.

However, in a TRS, the protection seller's payments will not stop there. The protection seller will also make an additional payment linked to Libor plus a spread.

The protection buyer, on the other hand will make periodic payments associated with the appreciation and the coupon of the underlying asset. Normally, asset prices appreciate and pay coupon more often then decline, but this is compensated by the Libor plus any spread received.

4.1 Equivalence to Funded Position

The TRS structure is equivalent to the following operation. A market participant buys an asset, S_t, and funds this purchase with a Libor-based loan. The loan carries interest rate, L_t, and has to be rolled over at each t_0. The market participant is rated A and has to pay the credit spread d_{t0} known at time t_0. The S_t has periodic (coupon) payouts equal to c. The market participant's net receipt at time t_{i+1} would, then, be the following Equation)

Where the ΔS_{ti+1} is the appreciation or depreciation of the asset price during the period $\Delta = (t_i, t_{i+1})$. The c is paid during Δ. The payment are in-arrears.

A TRS swap is equivalent to this purchase of a risky asset with Libor funding. Except, in this particular case, instead of going ahead with the transaction, the market participation can simply sign a TRS with a proper counterparty. This will make him or her a protection seller. Banks may prefer these types of TRS contracts to lending to market practitioners. Below, we provide an example that illustrate the point.

Example

Total return swaps should share the spotlight with credit default swaps in accounting for the record volume of synthetic collateralized debt obligations (CDOs) in 2001, according to traders and credit derivative strategies.

Total return swap are often overlooked compared with default swaps and credit-linked notes as a credit derivatives tool, but participants at an IMN CDO conference in New York last week touted the use of the instruments to transfer the credit and market risk associated with an underlying asset.

Total return swaps are similar to credit default swaps except that the protection seller makes periodic payments to the protection buyer and both seek risk protection. The seller's payments are usually based on Libor and are in addition to paying depreciation on the market value of the underlying to compensate the buyer for funding costs. In turn, the protection buyer makes payments to the protection seller consisting of the coupons and interest from the underlying asset, as well as any appreciation in the market value of the asset.

A quick turn around time for the execution of the swaps versus cash arbitrage CDOs has been boosting total return swap use lately. "In our typical CDO transaction in the assets would be financed with a trust. Total return swaps have been used where an institution will acquire the asset, whether on-balance sheet or off-balance sheet, and will pass the total return swap n to a group of investors, " said a principal at a bank." A total return of swap execution might take one or two months, whereas the execution of cash arbitrageCDO can take much longer", he said (IFR, February 2002).

Total Return Swaps can be combined be exchanged many different type of risky streams of cash flows. The discussion below involves one well-known case observed during the Asian crisis of the year 1998.

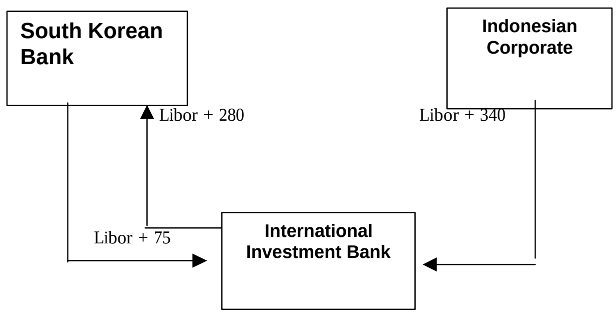
4.2 Another Example

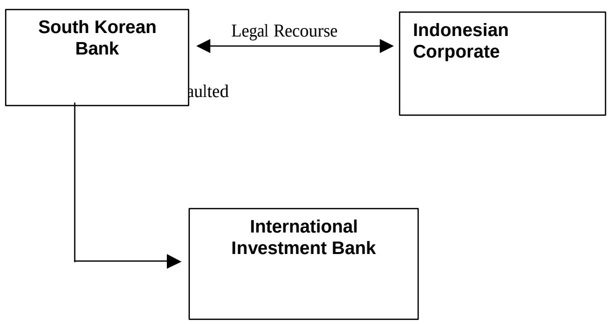
A total Return Swap (TRS) is shown in Figure 16.6 W see that the derivative consists of the exchange of two very different types of cash flows. First, where is an exchange of cash flows that carries of credit spread.

Example

A client swaps a libor +280 by return originating from a loan to, say, an Indonesian corporate against a Libor +75 bp return received from a South Korean bank. Libors are in the same currency. This exchange of cash flows is made regularly every six months. Of course, in reality. Only the net difference is paid at every settlement date.

The second exchange in the TRS involves any capital gains or losses that the underlying assets generate during the year. In particular, in the preceding example, in the event of bankruptcy of the Indonesian corporate, the South Korean bank would compensate the international bank for





The loss. In exchange for a higher predicate annual return, the South Korean bank is selling default protection to the international bank. This shown in Figure 16-7.

5. Uses of Credit Derivatives

There are many uses of credit derivatives. We briefly discuss some of these in this section. Our discussion will be driven by financial market examples.

1. Credit derivatives can be used to create synthetics for corporate and sovereign debt securities.
2. They are useful for manage risk in balance credit risk in balance sheets. The credit portfolio of banks can be effiency managed by buying and selling credit default swaps.
3. They are tools for changing the funding cost for long positions.
4. They provide possibilities of yield enhancement.
5. They have an important role to play in tax planning.
6. They are essential for regulations and capital adequacy.
7. They provide leverage for investors desiring exposure to various credits.

The best way to illustrate some of these fundamental contributions of credit derivatives is to discuss some major example.

5.1 Use of CDSs

We begin by looking some example of how CDSs are used. The example that follows is the first step. Since it shows how the market has developed.

Example

According to a market participant" the credit default swaps (CDS) market has really come into its own over past months, especially since the downturn in the credit cycle and derivatives makes it easier to see the hole picture.

According to its last published survey derivatives in July 2000, the British Bankers Association (BBA) estimated the global market size to be US $ 1, 581 bn by 2002, a nine fold increase in five years. Within that, it estimated that the default swap market makes up 37% of the total and that London constitutes 47% of the overall figure. This percentage is in the with bankers, best guesses one analyst suggested that if its own numbers are representative, the European market its similar in size to the US. (IFGR Issue 1430 April 2002).

The following is a specific example from the CDS market Argentina, World Com, and Enron are all interesting names to be associated with the CDS market because of the large size of the respective defaults. This example deals with Argentina, where the CDS rate was around 40 % for one year around the default period.

Example

One year Argentina credit default swap mid-level hit 4,000 bp late last week, through the highest trade in the sovereign is thought to have been year deal at 2,350 bp early in the week.

Derivatives market-makers were cautiously quoting swap prices on an extremely wide bid/offer spread (the two year Argentina mid rose to around 3,900 bp), but mostly concentrated on balancing cash market hedges, which did not prove easy.

Dealers who have sold protection also consulted their lawyers to plot tactics in the event that Argentina defaults, or restructures its debt. It is likely that more than US$1 bn of credit default protection on Argentina has traded in the last few years, which could result in the biggest default swap payout yet, if there is a clear-cut default or debt restructuring. There is plenty of scope for disagreement on whether or not the payout terms of swaps have been met, however, depending on how any debt restructuring is handled by the Argentine authorities.

Pricing default swaps when a payout trigger could be hours away is an art, not a science. Late last week traders were working from the closing price on Thursday of Argentina's FRBs of 63.5 which was the equivalent of 3,060 bp over Libor, then adding a 30-40 % basis for the theoretical risk of writing a default swap, as opposed to the asset swap value of a bond trade. For much of this year, traders have been using a default against asset swap basis of around 10 % of the total spread for deals in Latin American sovereigns. (IFR July 2001).

5.2 Structured Products

The recent trend in credit markets is towards structured products. The following example deals with this issue.

Example

European credit makers are making an evolutionary leap from the plain vanilla to the highly structured, apparently bypassing all points in-between. Institutional investors unwilling to buy high beta credits, such as high corporates or emerging market sovereigns, much less leveraged loans, seem happy to take chunky positions in highly structured assets tailored to their portfolio needs.

The significance of the tailor made asset is difficult to overstate. Many European investors. Regardless of their country of incorporation, are constructed by restrictions pertaining to denomination (local currency assets), listing, swap exposure and rating requirements.

For example, a fund manager restricted by mandate to buying euro-denominated, listed bonds from an EU issuer cannot take advantage of cheep Yankee paper. But the investor can buy euro-dominated repacked notes backed by subordinated debt, issued in the Yankee market and execute a US dollar/euro swap with the note seller. The Special Purpose Vehicle (SPV) that issues the notes can be listed in the EU and the notes listed in Luxembourg, allowing the securities to comply fully with the investor's risk management and legal criteria.

An investor may want to take advantage of the basis between cash bonds and default swap, but cannot enter into derivatives transactions. This, combined with the needs for one or two credit ratings from Moody's and S & P would encourage the buyer to consider the medium-term note sector. But if MTN issuers are expensive or will not link there name to all of the credits in which the investor has an interest, a repacked credit-linked note will serve the same purpose.

The issuer (SPV) purchases cheap secondary paper and then enters into a credit linked swap with the underwriter, such as Deutsche Bank. The SPV notes, carry the appropriate ratings and are linked to a basket of names. (IFR, Issue , 1290, July 1999).

The reading also illustrate the importance of legal and regulatory issues in pricing credit instruments. Due to various restrictions on holding some credits, institutional investors use derivatives and swaps to gain exposure to the same credits. This leads to a difference between, say CDS rate and the equivalent benchmark spread over swaps.

5.3 Another Use

We discussed CDSs and other credit derivatives as if they were written on outstanding bonds. However, these instruments are stand-alone derivatives except that the contract needs to specify clearly what default means. Also, if the default occurs, the protection payment will be exchanged against the underlying bond, which would be the recovery value. The protection seller will make the agreed payment and receive the defaulted bond.

As the simple engineering case showed, a CDS can be written on any obligation that makes periodic interest payments. Another example is syndicated loans. These have different legal characteristics from those of bonds, but the basic idea remains the same. The following reading illustrate this general nature of CDS-type credit instruments.

Example

Credit derivatives look set to becomes an integral tool in the primary loan syndication market. Bankers say that credit derivatives are being used both as a tool to pitch for mandates and to sell down loans during general syndication. Savvv investors are beginning to by-pass general syndication to make arbitrage gains by selling credit cover instead of taking loan assets.

The benefits to arrangers in syndicates that wish to off load poorly paying assets are enormous. Secondary loan sales have the disadvantage of requiring borrower consent, while CLO programs are time consuming. With the credit-default swap, lenders keep their relationship and ancillary business but free-up capital.

Writing credit-default swaps for loans does offer problems unique to the loan market. Whereas secondary loan traders require assignment permission from the borrower, credit derivatives are transacted without the knowledge of the loan borrower. This could prove tricky should the company default which after all, is the premise for the trade in the first place for holders of the credit will have to admit to the trade. There are legal ramifications should failure be granted to honor new creditors. Also, loans inherently features prepayment risk. Credit agreement is also open to changes in covenants.

The reluctance by corporate treasures to agree to loan transferability could actually act as a boon to the superstitions credit-default market in Europe. (IFR, Issue, 1290, July 1999).

This reading shows that CDS instruments can be used to rearrange the credit portfolios of banks. As loan in general are not transferable, and as banks would like to preserve their client relationship, the best way to eliminate the exposure to a credit is to buy a CDS on the name. Yet, as the example illustrates, a loan on the book of the bank is not, in general, a deliverable security in case of default. Hence, a bank that hedges a credit with a CDS may face a significant basis risk.

5.4 Uses of Total Return Swaps

We now consider an example from financial markets to see how market participants use TRS instruments.

Example

Fund managers seeking to build up credit portfolios are taking a hard look at total return swaps based on indices of European corporate bonds. Driving their interest is the spectre of European Monetary Union and the predicated shift away from interest rate and currency-based portfolio optimization to credit based yield enhancement.

Bond index total return swaps are derivatives that provide bond index returns in exchange for Libor-related payments. On their own, they expose the holder to both interest rate and credit risk. Investors can cut out the interest rate risk by holding a short position in either a portfolio of government bonds or a government bond total return swap. The leaves them with pure credit exposure.

Index total return trades are attractive to banks because they can be used as an effective hedge for their large bond trading books. Some banks use these swaps to offset the exposure on the bank's floating-rate bond book – an application that helps it offer very more competitive quotes to asset managers. For other credit derivative professionals, total return swaps are the ideal product for European fund manager looking to build up a new credit risk portfolio. They provide the investor with exposure to 500 or 1,000 names and, thus, a ready-made diversified portfolio. A portfolio built by picking individual securities, said professionals, would suffer from high correlation risk until it became larger and more diversified.

The other factor driving interest in index total return swaps is the lack of high-yield cash product available. High yield bond issuance in 1997 was roughly US$1 bn, and is only forecast to rise between US$1 bn and US$10 bn in 1998 – minuscule in comparison with the investment grade market. High yield professionals said the market was very illiquid and bid/offer spreads were extremely wide, "You just cannot buy mid-market grade debt in Europe. For funds looking to earn higher yields, derivatives are only option". Said one market professional in London.

Total Return Swap can also be regarded as funding instruments. A certain institution may desire to fund a long position in an asset using borrowed funds. A loan may be obtained and the asset may be purchased with these funds. TRS is an alternative way of providing similar opportunities to this institution. (IFR, April 1998).

As an example to this use of TRSs, consider the following arbitrage. A bank has a funding cost of 6 %. The bank would like to buy a (less risky) asset with a return of 5%. If the asset is funded at market rates, it leads to negative carry.

The bank can use a TRS to receive the 5% return. This will be funded at the swap conterparty's funding cost, and the bank will pay Libor plus a spread, which could be less than the original 100 bp difference. According to this example, TRSs, are used to arbitrage funding costs.

6. Balance Sheets and Credit Derivatives

Credit risk portfolios carried on bank balance sheets require constant management. The previous example discussed one aspect of hedging credit risk exposures of loans. However, many regulatory issues are involved in managing credit risk portfolios. The following example summarizes these difficulties. It also illustrates the use of structured credit products that are becoming more common.

Example

JP Morgan has bridged the gap between economic hedging and efficient capital management, according to credit derivatives market participations. The US bank is rumored to have combined securitization and credit derivatives technology to break through the firewall erected by the 1988 Basel Capital Accord between hedging and regulatory capital requirements.

The new portfolio management structure is the first to manage capital and economic risk efficiently. A bank issuing a Collateralized Loan Obligation (CLO) typically retains the first loss equity stump at the base of the deal's capital structure. The motivation for CLO issuance has effectively been to arbitrage the regulatory capital cost of holding the equity against that of holding the underlying loan assets themselves.

By contrast, credit derivatives while a more effective form of economic hedging, have so far secured little regulatory capital relief. The Basel Accord fails to recognize the portfolio effects of credit hedging. Credit hedges only result in a reduction in aggregate risk capital charges where the synthetic short position through the credit derivatives exactly offsets that of the long cash position held on the bank's balance sheet.

The hedges must be of the same maturity and must be referenced to the same asset. Hence, a six-year loan exposure to a single A-rated corporate, hedged with a three-year credit default put option, receives no reduction in risk weighting despite the mitigation of credit risk involved.

In the Bistro transaction, a JP Morgan owned special purpose vehicle entered into a US$9 722 bn credit default swap with JP Morgan. The swap was referenced to a portfolio of commercial loans, corporate and municipal bonds, and to counterparty credit exposures arising out of derivatives contracts – a virtual representation of JP Morgan's credit exposures during a business cycle. The special purpose vehicle then sold US$700 m of bonds to a broad base of investors, with the performance of the bonds referenced to the performance of the credit default swap. Through Bistro, JP Morgan successfully involved genuine risk transfer – a reduction in the credit quality of securities on its banking book would be matched by a reduction in its liabilities to the bondholders. (IFR, May 1998).

Regulatory arbitrage is an area of its own. But it is also in a transition period due to the new BIS credit accord which is to be put in place in the near future.

ENGINEERING OF EQUITY INSTRUMENTS

Objectives

- After completion of this lesson you will be able to understand some basic equity and securitization products.

Fixed income instruments involve payoffs that are, in general, known and "fixed" They also have set maturity dates. Putting aside the credit quality of the instrument, fixed-income assets have relatively simple cash flows that depends on a known, small set of variables and, hence, risk factors. There are also well-established and quite accurate ways to calculate the relevant term structure. Finally, there are several liquid and efficient fixed-income derivatives markets such as swaps, forward rate agreements (FRAs), and futures, which simplify the replication and pricing problems existing in this sector.

There is no such luxury in equity anylysis. The underlying asset, which is often a stock or a stock index, does not have a set maturity date. It depends on a nontransparent, idiosyncratic set of risks, and the resulting cash flows are complex issues of growth, investment, and management decisions that further complicate the replication and pricing of equity instruments. Finally, relatively few related derivatives markets are liquid and usable for a replication exercise.

Yet, the general principles of pricing, replicating, and risk-managing equity cannot be that different. Whatever is possible in the fixed-income sector should in principle, be possible in equity as well. Of course, the approximation process, and the resulting modeling of these instruments may become more difficult and the success rate of the potential methods may drop.

In this chapter, we extend the methods introduced earlier to equity and equity-linked products. Second, we discuss the engineering applications of some products that are representative of the sector.

Our intention is to show how the methods used in fixed income can, in principle, be used in the equity sector as well. In doing this, we will analyze the major differences and some similarities between the two sectors. There are two additional difficulties with equity. First, equity analysis may require a modeling effort to project the underlying earnings. This is because the implied cash flows of a stock are never known exactly and are difficult to predict. 2 Financial engineering methods that use the fundamental theorem of asset pricing avoid this issue by replacing true "expected returns" with the risk-free return. Yet, this cannot always be done. For some exercises, future cash flows implied by stocks need to be projected using real-world probabilities.

This chapter also introduces financial engineering applications that relate to asset-backed securities (ABS) and securitization. It turns out that securitization and hybrid asset creation are similar procedures with different objectives. From the issuer's point of view, one is a solution to balance-sheet problems and it helps to reduce funding costs. From an investor's point of view, securitization gives access to payoffs the investor had no access to before, and provides opportunities for better diversification. Hybrid assets, on the other hand, cab be regarded as complex, ready-made portfolios.

A financial engineer needs to know how to construct an ABS. In fact, engineering is implicit in this asset class. The remaining tasks of pricing and risk managing are straightforward. A similar statement cab be made about hybrid assets. We begin the chapter by reviewing the basics of equity instruments and by adapting the tools we have seen thus far to this sector.

1. What Is Equity ?

Bonds are contracts that promise the delivery of knows cash flows, at known dates. Sometimes these cash flows are floating, but the dates are almost always known, and with floating-rate instruments, pricing and risk management is less of an issue. Finally, the owner of a bond is a lender to the institution that issues bond. This means a certain set of covenants would exist.

Stocks, on the other hand, entitle the holder to some ownership of the company that issues the instrument. 3 Thus, the position of the equity holder is similar to that of a partner of the company, benefiting directly from increasing profits and getting hurt by losses. In principle, the people selected by stockholders manage the corporation. The equity should then be regarded as a tradable security where the underlying cash flows are future earnings of the corporation.

1.1 A Comparison of Approaches

The best way to begin discussing the engineering of equity-based instruments is to review the valuation problem of a simple, fixed-income instrument and simultaneously tries to duplicate the same steps for equity. The resulting comparison clarifies the differences and indicates how new methods cab be put together for use in the equity sector.

Consider first the cash flows and the parameters associated with a three-period coupon bond P (t_0, T), Shown in Figure 17-1. The bond is to be sold at time t_0 and pays coupon c three times during { t_1, t_2, t_3}. The date t_3 is also the maturity date denoted By T. The par value of the bond is $ 100, and there is no default risk.

2. For example, what is the value of the earnings of a company? Analyses that depart from the same generally-accepted accounting principles often disagree often disagree on the exact number.
3. Not all stocks are like this. There is Euro-equity, where the asset belongs to the bearer of the security and is not registered. In this case., the owner is anonymous, and, hence, it is difficult to speak of an owner, Yet, the owner still has access to the cash flows earned by the company, although he or she has no voting rights and ,hence, cannot influence how

the company should be run. This justifies the claim that the Euro-stock owner is not a "real" owner of the company.

Figure 1

Next, consider the stock of a publicly traded company denoted by St. Let Zt be a process that represents the relevant index for the market where St trades. The corporation has future earning per share denoted by et.

We will bow try to synthetically recreate these two instruments. one fixed income, the other equity. The purpose is to show how the pricing of an equity instrument differs from the rather simple solution in the case of fixed income. The basic principle that should apply to both asset classes is that the value of a security at time t_0 should equal the discounted value of expected cash flows from the contract. However, this statement is vague and needs to be made more precise.

Suppose there are P(t_0, T) dollars to invest. Consider first a savings deposit. Investing this sum in the short-term spot rate, Lt, instead of the coupon bond, will yield the sum :

Equations P ($t_{0,}$, T) (1 +)

in three periods at time t_3 . Here, is the usual adjustment for the day count and {L $t_{0,\ L}$ $t_{1,}$ L t_2} are the short term rates that will be observed at times $t_{0,}$ t_1 and $t_{2,}$ respectively.

A second possibility is the purchase of the default free bond P ($t_{0,}$ T). This will result in the receipt of three coupon payments and the payment of the principal. Finally, we cab buy k units of the stock St.

The simplest approach to price or risk manage the bond portfolio would be to proceed along a line such as the following. The coupon bond that pay c three times is equivalent to a properly chosen portfolio of zero-coupon bond:

Portfolio = Equation

where B { $t_{0,}$ t_1 } are default-free, zero-coupon bonds that mature at dates $t_{i.}$ Clearly, this portfolio results n the same case flow as the original coupon bond P ($t_{0,}$T). Given that the two investments are assumed to have no credit risk or any other cash flows, their value must be the same:

Equation

Equation

But, we know that the arbitrage-free prices of the zero-coupon bonds are given by

Equation

Equation

We obtain the valuation equation that uses risk-neutral probability P, with random L t_1 and L t_2 at time t_0 :

Equation

Equation

Here, L t_1 and L t_2 are random variables distributed with probability P.

We are not yet done with this equation since it involves an expectation operator and is therefore only a representation and not an operational formula. But, we should stop here and consider how the derivation up to this point would be different in the case of equity.

2.2 The Case of Stocks

In the following, we try to apply the same methodology to price a stock. We assume the following:

- The stock does not pay dividends.
- There are no other corporate actions such as stock splits, capital injections, or secondary issues.
- There exists a market stock index calculated using all the traded stocks in this market.

We can buy unit of St to get the title for future earning {et_1 }. Following the same steps, we need to do two things. First, we find a synthetic for the stock using other liquid and possibly elementary securities, and then equate their price. Suppose we put together the following portfolio:

Equation

Equation

and then we proceed similarly to pricing the bond. There are at least two potential problems with this method. First, the dollars that the company promises to pay through future earning e_{ti} . and the dollars promised by the maturing zero-coupon bond B ($e_{0,}$ ti) may not be an appropriate present value for $e_{t.}$ Of course, assuming (unrealistically) that ther is no credit risk eliminates this problem. But, a second problem remains. Unlike in the case of a soupon bond where the coupon payments c were constant and gave constant weights in the replicationg portfolio, the future earning e_{ti} are random. So the weights of the portfolio in (8) are not known and, thus, the portfolio itself cannot be a replicating portfolio. This means that in the case of equity the logic is not the same.

One way to look at it is to ask the following questions: Can we modify the approach used for the fixed instrument a little and employ a method that is similar? In fact, by imposing some further (restrictive) assumptions, we cab get a meaningful answer. The one-factor version of this approach is equivalent to the application of the so-called *CAPM* theory.

This book is not the place to discuss the capital asset pricing model (CAPM) , but a fairly simple description that illustrates the parallels with the case of derivatives and fixed income will still be given. The idea goes as follows. Suppose Zt is a correct stock index for the market where St trades. Assume that we have the following (disorganized) risk-neutral dynamics for the pair St, Zt :

Equation

Equation

Where Δ Zt, Δ St are in increments in the Zt, St variables and the r is the constant risk-free rate. Δ Wst amd Δ Wmt are two independent increments of the corresponding Wiener processes that need to be discussed further.

We assume that Δ Wst is a risk that is diversifiable and specificto the single stock St only. The market index is affected only by Δ Wmt. This represents a risk that is nondiversifiable. It has be borne by stock holders. Thus, this is a model with two factors, but one of the factors is not a true risk, although it is a true source of fluctuation in the stock St.

To obtain a formula similar to the bond pricing representation, we postulate that, expected future earning properly discounted shoud equal the current price St. We then use the real-world probability P and the real world discount rate dt that apply to the dollars earned by this company to write an equation that corresponds to the representation for the coupon bond price: [4]

Equation ————————————————————

Equation ————————————————————

It is worth emphasizing that, in this expression , we are using the real-world probability. Thus, the relevant discount rate will differ from the risk free rate:

Equation ————————————————————

Equation ————————————————————

We need to discuss how such a dt cab be obtained.

To do this, we need to use the following economic equilibrium condition: If a risk is diversifiable, then in equilibrium it has a zero price. The market does not have to compensate an investor who holds a diversifiable risk by offering a positive risk premium. We use this in the section follows.

1.1.1. Beta

The only source of risk that the investor needs to be compensated for is Wmt. But, if this is the case, and if Wst risk cab be considered as having zero price, then we cab use Zt as a hedge to eliminate the movements in St cause by Wmt only. There are two ways we can look at this.

Equation ————————————————————

Equation ————————————————————

and then substitute the right-hand side in

Equation ————————————————————

Equation ————————————————————

Dividing by am St and rearranging:

Equation ————————————————————

Equation ————————————————————

Since the first term on the right is diversifiable by taking expectations with respect to the real world probability we cab writ this using the corresponding expected (annual) returns. R^s_t . and R^m_t

Equation ————————————————————

Equation ————————————————————

Now, from pricing perspecive, market price of a diversifiable risk is zero. This implies that there is a single factor that matters. Accordingly, we posit the following relationship involving a_m :

Equation ————————————————————

Equation ————————————————————

Then, we can substitute this in equation (16) to obtain a formula that gives a discount factor for the equity earnings.[5]

Equation ————————————————————

Equation ————————————————————

If we are given the right hand side values, we cab calculate the R^s_t and use it as a discount factor in

Equation ————————————————————

Equation ————————————————————

Again, this is a representation only and not a usable formula yet. Next, we show how to get usable formulas for the two cases.

1.2 Analytical Formulas

How do we get operational formulas from the representation in Equation (7) and (19), respectively? in the case of fix income, the answer is relatively easy, but for equity, further work is needed.

To convert the bond representation into an operational formula, we cab use two liquid FRA contracts as shown in Figure 17-2. These contracts show that markdet participants are willing to pay the known cash flow $F(t_0, t_1)$ against the unknown (at time t_0) cash flow L t_1 and that they are willing to pay the known F (t_0, t_2) against the random L t_2. Thus, any risk premia or other calculations concerning the random payments L t_1 and L t_2 can be “replaced” by $F(t_0, t_1)$ and $F(t_0, t_2)$, since the latter are equivalent in value as shown by the FRA contracts.

Figure————————————————————

Figure————————————————————

Figure————————————————————

This implies that, in the formula

Equation ————————————————————

Equation ————————————————————

no-arbitrage condition will permit us to “replace” the random Lt_1 and Lt_2 by the known $F(t_0, t_1)$ and $F(t_0, t_2)$. We then have

Equation ————————————————————

Equation ————————————————————

This is the bond-pricing equation obtained through the risk neutral pricing approach. Note that to use this formula, all we need is to get the latest sopt and forward Libor rates Lt_0, $F(t_0, t_1)$, $F(t_0, t_2)$ from the markets and then substitute.

Obtaining an analytical formula in the case of equity is not as easy and requires further assumptions beyond the ones already made. Thus, starting with the original representation:

Equation ————————————————————

Equation ————————————————————

To convert this into a usable formula, the following set of assumptions is needed.

There are an infinite of e_{t+i} in the numerator, First, we need to truncate this at some large but finite n. Then, assume that the company earning will grow at an estimated future rate of g, so that we cab write for all i.

Equation ————————————————————

Equation ————————————————————

Finally , using some econometric or judgmental method, we need to estimate the craning per share, et After estimating the et, Beta we cab let.

Equation ——————————————————

Equation ——————————————————

This equation cab be used to value St. It turns out that most equity analysts use some version of this logic to value stocks. The number of underlying assumption is more than in the case of fixed income, and they are stronger.

Summary

Let us summarize. The valuation of the fixed-income instrument is simple for the following reasons:

1. Given that the coupon rate e is known, we cab easily find a replicating portfolio using appropriate zero-coupon bonds where the weights depend on the coupon.
2. The maturity of the bond is known and is finite so that we have a known, finite number of instruments to replicate the bond with.
3. The existence of FRA contracts permits "replacing" the unknown random variable with market equivalent dollar quantities that are known and exact.

The valuation of equity requires further restrictions.

1. A model for the market return something needs to be adopted. This is the modeling component.
2. The number of factors needs to be specified explicitly in this model.
3. Economic equilibrium need to be invoked to claim that diversifiable risks won't be rewarded by the markets, and that the only volatility that "matters" is the volatility of no diversifiable risks.

After this brief conceptual review, we cab now consider some examples of equity products.

3. Engineering Equity Products

The second purpose of this chapter is to discuss the engineering of some popular equity instruments.

A large class of synthetic securities has been created using equity products, and the popularity of such instruments deeps increasing. This is not the place to discuss the details of these large asset classes. Yet, they provide convenient examples of how financial engineering cab be used to meet various objectives and to structure hybrid equity products. The discussion here is not comprehensive. At the end of the chapter, we provide some additional references.

The plan of this section is as follows, we provide by considering the earliest and best-known equity-linked instruments. Namely, we discuss convertible bonds and their relative, warrant linked bonds. These engineering issues in the equity-linked sector.

Then, we move to index-linked products, which are a more recent variant. Hence, even though the general structures are not much different, the synthetic are constructed for different purposes, using equity indices instead of individual stocks which is the case in convertibles and warrant-linked securities.

The third group is composed of the more recent hybrid securities that hve a wider area of application.

3.1 Purpose

Companies raise capital by issuing debt or equity.[6] suppose a corporation or a bank decides to raise funds by issuing equity. Are there advantageous ways of doing this? It turns out that the company cab directly sell equity and raise funds. But, the company may have specific needs. Financial engineering offers several alternatives.

1. Some strategies may decrease the cost of equity financing.
2. Other strategies may result in modifying the composition of the balance sheet.
3. There are steps directed toward better timing for issung securties depending on the direction of interest rates, stock markets, and currencies.
4. Finally, there are strategies directed toward broadening the investor base.

In discussing these strategies, we consider three basic instruments that the reader is already familiar with. First, we need a straight coupon bond issued by the corporate. The case flows from this instrument are shown in Figure 17-3a [7] . The bond is assumed to have zero probability of default so that the cash flows are known exactly. The coupon is fixed at c, and the bond is sold at par, so that the initial price is $100.

The second instrument is a dividend-paying stock. The initial price is St and the dividends are random. The company never goes bankrupt. The cash flow are shown in Figure 17-3b. The third instrument is an option written on the stock. The (call) option on the stock is of European style. has expiration date T, and strike price K. The call is sold at a premium $C(t_0)$. Its payoff at time T is.

Equation ——————————————————

Equation ——————————————————

Equation ——————————————————

These sets of instruments will be complemented by two additional products. In some equity-linked products, we may want to use a call option on the bond as well. The option will be European. In other special cases , we may want to add a credit default swap to the analysis. Many useful synthetics cab be created from these building blocks. We start with the engineering of a convertible bond in a simplified setting.

3.2 Convertibles

A convertible is a bond that incorporates an option. At the maturity date, the principal can be paid as a predetermined number of stocks of the issuing company, if the bondholder so desires. Otherwise, the par value is received. In other words, the owner of the bond has purchased the right to "convert". It is clear that the convertible bond is a hybrid product that gives the bondholder exposure to the company stock in case the underlying equity appreciates significantly. We discuss the engineering of such a convertible bond under simplified assumptions. In the first case we discuss a bond that has no default risk. This is illustrative, but unrealistic. All Corporate bonds have some default risk. Sometimes this risk is significant.

Hence, we redo the engineering, after adding a default risk in the second example.

Figure——————————

Figure——————————

3.2.1 Case 1 : Convertible with No Default Riak

Suppose a default-free bond pays $ 100 at maturity and consider the following portfolio:

Protolio = {1 Bond, long n call options on the stock with nK = 100} (26)

This portfolio of a bond and n call options is shown in Figure 17-4. Consider the top part, of the figure. Here, the holder of the portfolio is paying for the bond and receiving three coupon payments. At $T = t_3$ .the bond holder also receives the principal. This is the cash flow of a typical default-free coupon bond.

Figure——————————

Figure——————————

The second cash flow show what happens if the option ends up in the money. n such options are bought, so, initially, the portfolio holder pays nC (t_0) dollars for the options. Given that these options are European, there is no other cash flow until expiration. At expiration, if the option is in the money, the bond will convert and the payoff will be,

Equation..........................

Equation..........................

This can be regarded as an exchange of n stocks, each valued at S_T, against the cash amount nK. But n is selected such that nK =10. Thus, it is as if the portfolio holder is receiving n shares valued at S_T, each and paying $100 for them. This is exactly what a plain vanilla call option will do when it is n the money. But, in this case, ther is the additional convenience of $100 being received from the payment of the principal n the bond.

Putting these two cash flows together , we see that the portfolio holder will pay $100+nC(t_0)$, receive c dollars at every coupon payment date until maturity, and then will end up with n shares valued at S_T each, if the option expires in the money. Otherwise, the bondholder ends up with the principal of $100. When option expire out of the money, there will be no additional cash flows originating from option expiration. This case is equivalent to a purchase of a coupon bond. The coupon c is paid by a bond that initially sold at 100 + nC (t_0). Because this is above the par value 100 on issue date, the yield to maturity of this bond will be less than c. Using the internal rate of return representation for the par yield y can see this:

Equation ....

Equation ....

We need to have y < c as long as nC $(t_0)>0$.

This discussion shows what convertible is and suggests a way to price it if there is no default risk: A convertible bond is a bond purchased at an "expensive" price if St<K-that is to say, if the stock price fails to increase beyond the strike level K. In this case, we say that the bond fails to convert. But if at expiration St > K, the bond will give its holder n shares valued at S_T with a total value greater than $100, the principal that a typical bond pays. The bond converts to n shares with a higher value than the principal. In order to price the convertible bonds in this simplistic case, we first price the components separately and then add the values.

3.2.2 Case 2: Adding Default Risk

The decomposition of the convertible bond discussed above is incomplete in one major respect. To simplify the discussion in the previous section, we assumed that the convertible bond is issued by a corporation with no default-risk. This is clearly unrealistic since all corporate bonds have some associated credit risk.

Example

Convertible arb hedge finds in the U.S. are piling into the credit default swaps market. The step up in demand is in response to the rise in investment grade convertible bond issuance over the last month, coupled with liquidity in the U.S. asset swaps market and the increasing credit sensitivity of convertible players' portfolios, said market offcials in New York and Connecticut.

Arb hedge funds are using credit default swaps to strip out the credit risk from convertible bonds, leaving them with only the implicit equity derivative and interest-rate risk. The latter is often hedged through futures or treasuries. Depending on the price of the investment-grade convertible bond. This strategy is often cheaper then buying equity derivatives option outright, said [a trader].

Asset swapping, which involves stripping out the equity derivative from the convertible, is the optimal hedge for these funds, said the [trader] as it allows them to finance the position cheaply. and removes interest-rate risk and credit risk in on fell swoop. But with issuer-credit quality in the U.S. ove the last 12 to 18 months declining, finding counter parties willing to take the other side of an asset swap has become more difficult.......(Based on an article in Derivatives Week)

It is clear from this reading that arbitrage strategies involving conversable bonds need to consider some credit instrument such as credit default swaps as one the constituents. We now discuss the engineering of convertible bonds that contain credit risk. This will isolate the CDS implicit in these instruments.

3.1.2. Engineering Default Table Convertibles

In the decomposition of a convertible discussed earlier, one of the constituents of the convertible bond was a straight coupon bond with no default risk. We now make two new assumptions:

- The convertible bond has credit risk.
- Without much loss of generality , the bond converts (i.e., $S_T > K$) only if the company does not default on the bond.

Figure....................

Figure....................

Then, the engineering of this convertible bond cab be done as shown in Figure 17-5. In this figure we consider gain a three-period risky bond for simplicity. The bond itself is equivalent to a portfolio of a receiver swap, a deposit, and CDS. Thus, this time the implicit straight bond is not default free.

Figure 17-5 shows how we cab decompose the risky bond as discussed in Chapter 16. According to this , now introduce an interest rate swap and a credit default swap. The horizontal sum of the cash flow shown in this figure result in exactly the same cash flows as the convertible bond with credit risk once we add the option on the stock. The resulting synthetic leads to the following contractual equation:

Figure...................

Figure...................

This contractual equation shows that if a market practitioner wants to isolate the call option on the stock that is implicit in the convertible bond. then he or she needs to (1) take a position in a payer swap, (2) buy protection for default through CDS, (3) get a loan with variable Libor rates, and (4) buy the convertible. In fact this is essentially , what the previous reading suggested.

3.3 Important Variations

This section considers two variations of this basic convertible structure. First all, the basic convertible cab be modified in a way that will make the buyer operate in two different currencies. In fact, a dollar –denominated bond may be sold, but the underlying shares may be, say. French shares, denominated in Euros. This amounts, as we will see, to adding a call or put option on a foreign currency. This is an interesting alternative.

The second alteration is also important. The basic convertible cab be made callable. This amounts to making the underlying debt issue a callable bond. It leads to adding a call option on the bond. This also may have some interesting implications. Before we see how these are used. we consider some of the financial engineering issues in each case.

1.1.1 Exchanges Rate Exposure

Suppose the convertible bond is structured in two currencies. A Thai company secures funding by selling a Euro convertible in the Eurodollar market, and the debt component of the structure is denominated in dollars. So, the bonds have a par value of , say $ 100. The conversion is into the shares of the firm, which trade, say, in Bangkok. the shares are baht denominated. We assume, unrealistically, that there is no default risk.

Because Thai shares trade in Thai exchanges and are quoted in Thai baht, the conversion price to be included in the convertible bond needs to specify something about the value of the exchange rate to be used during a potential conversion. Otherwise, the conversion rule will not be complete. That is to say, instead of specifying only the number of shares, n , and the conversion price, K , using the equality.

$$100\$ = Kn \qquad (30)$$

the conversion condition now needs to be

Equation ....

where it is an exchange rate denoting the price of on USE in terms of Thai baht at date t. This is needed since the original conversion price, K, will be in Thai bhat, yet , the face value of the bond will be in USD. The bond structure cab set a value for et and include it as a parameter in the contract. Often, this et will be the current exchange rate.

Now, suppose a Thai issuer has sold such a Euro convertible at et, the current exchange rate. Then, if Thai stock rise and the exchange rate remains stable, the conversion will occur. Here is the important point. With this structure, at maturity, the Thai firm will meet its obligations by using its own shares instead of retiring the original $100 to bondholders. Yet, if , in the meantime, et rises, then, in spite of higher stock prices, the value of the original principal $100, when measured in Thai baht, may still be higher then the nS_T and the conversion may not occur. As a result, the Thai firm may face a significant dollar cash outflow.[9]

This shows that a convertible bond, issued in major currencies but written on domestic stocks, will carry an FX exposure. This point cab be seen mor clearly if we reconstruct this type of convertible and create its synthetic. This is done in Figure 17-6.

Figure....................

Figure....................

Figure....................

This top part of Figure 17-6 is similar to Figure 17-4. straight coupon bond with coupon c matures at time t3 and pays the principal $100. The difference is in the second part of the figure. Here, we have, as usual, the call option on the stock, St. But St is denominated in baht and the call will be in the money –that is to say , the conversion will occur only if

$$nSt_3 > 100e\ t_3$$

The idea in Figure 17-6 is the following. We would like to begin with a dollar bond and then convert the new call option into an option as in the case before. But, if the Thai baht collapses, [10] then the $100 receives from the principal at maturity will be much more valuable than $St_3\, n/et_0$

1.2 Making the Convertibles Callable

One can extend the basic convertible structure in a second way, and add a call option on the underlying convertible bond. Fro example, if the bond maturity is T, then we cab add an implicit option that gives the issuer the right to buy the bond back at time, $U, U < T$ at price

$$\max[\$100, nSu]$$

This way the company has the right to force the conversion and issue new securities at time U. Some corporations may find this useful strategy.

With this type of convertible, forcing the conversion is the main purpose. Suppose the following two conditions are satisfied.

1. The share is trading at a higher price than the conversion price (i.e. the strike K).
2. The expected future dividends to be paid on the stock are lower than the current coupon of the convertible.

Then, if the convertible is callable, the issuer may force the conversion by calling the bond. This will convert a debt issue in the issuer's balance sheet into equity and affect some important ratios, in case these are relevant, Second, the immediate cash flow of the firm will improve.

1.3 More Complex Structures

The basic convertible- warrant structures cab be modified tomeet further financial engineering needs. We cab consider another example.

Suppose the convertible bond, when it converts, converts into another company's security. This may be the case , fro example , if company A has acquired an interest in company B. This way, the company cab sell convertible bonds where the conversion is into company B's securiti

From a financial engineering point of view, the structure of this "exchangeable" is the same. Yet, thepricing and risk management are different are deferent because now there are two credits that affect the price of the bond: the credit of the company that issues the bond and the credit fo the company this bond may convert.

Another difference involves the dilution of the shares of the target company. When a convertible is issued and converts at a later date, there may be dilution of the shares, yet, in an exchangeable the shares that are exchanged will come, in general, from the free float. [11]

1.4 Using Convertibles

A convertible bond has some attractiveness from the point of view of end investors. For example, the investor who buys the convertible will have some exposure to the share price. If St increases significantly, the bond becomes a portfolio of shares. On the other hand, if the bond fails to convert , the investor has at least some minimum cash flow to count on as income, and the principal is recovered (when there is no default).

But , our interest in this book is not with the investors, but rather, in the advantages of the product from an issuer's point of view. For what types of purposes cab we use a convertible bond?

- The first consequence of issuing convertibles rather then a staight bond is that the convertible carries a lower coupon. Hence, it "seems" like the funds are secured at lower cost.
- More notably for a financial engineer, convertibles have interesting implications for balance sheet management. If an equity-linked capital is regarded as equity, it may have less effect on ratios such as debt to equity. But, in general, rating agencies would consider straight convertibles as debt rather than equity.
- Note that with a convertible, in case conversion occurs, the shares will be sold at a higher price than the original stock price at issue time.
- Finally, convertibles are bonds. , and they cab be sold in the Euro markets as Euro- convertibles. This way a new investor base cab be reached.

We should also point out that convertibles, when combined with othe instruments, may have significant and subtle tax advantage. The best way to show this is by looking at an example from the markets.

Example

(ABC Capital) Ihas entered into a total return swap on 154,000 shares of Cox Communications preferred stock exchangeable into shares of Sprint PCS, and a total return swap an 225,000 shares of Sprint PCS. In the Cox swap, the hedge fund pays three month Libor plus 50 basis points and receives the return on the exchangeable preferred shares. In the Sprint swap, ABC, Pay the return on the stock and receives three-month Libor less 25bps. Both total return swaps mature in about 13 months.

The total teturn swaps were entered into for taxs reasons. A Cayman Islands Limited duration company holds ABC's positions. Because the Cayman Islands do not have a tax treaty with the U.S. income from these securities is withheld at the not treaty rate of 30%. Entering the total return swaps ensures that ABC does not physically hold the securities , and , hence , is not subject to U.S. withholding.

The underlying position was put on as part of a convertible are play. ABC bought the exchangeable preferred stock and is using the cahs equityto delta hedge the implicit equity option. The market is undervaluing the exchangeable preferred shares, according to a trader, who noted that although these shares recently traded at USD76.50 the fund's models indicate they should be priced around USD87. The company's model is based in part on the volatility of the underlying stock, the credit quality of the essuer, and the features of the convertible. In this case, the market may be undervaluing the security because it is not pricing in all the features of the complicated preferred and because of general malaise in the telecom sector. (Derivatives Week November 2000)

This reading is also an example of how implicit options cab be use to from arbitrage portfolios. However, there are many delicate points of doing this as were shown earlier.

1.5 Warrants

Warrants are detachable options to bonds. In this sense , they are similar to convertibles. But, from a financial engineering point of view, there are important differences.

1. The Warrants is detachable and can be sold separately from the bond. Of course , a financial engineer cab always detach the implicit option in a convertible bond as well, but still there are differences. The fact that the warrant is detachable means that the principal will always have to be paid at maturity.

   The number of warrants will not necessarily be chosen so as to give an exercise cash inflow that equals the cash outflow due to the payments of the principal. Thus, the investor can , in principle, end up with both the debt and the equity arising from the same issue.
2. The exchange rate used in a convertible is fixed. But, because ther is no such requirement for a warrant and because the latter is detachable, this is, in general, not the case for a warrant. Hence, there is no implicit option on the exchange rate in the case of warrants. In this sense warrants are said to be relatively more attractive for strong currency borrowers, whereas convertibles are more attractive for weak currency borrowers.
3. Finally, because warrants are detachable,, the warrant cannot be forced to convert. the an example of new product structuring.

4. Financial Engineering of Securitization

Every business or financial institution is associated with a "credit" or, more precisely, a credit rating. If this entity issues a debt instrument to secure funding, then the resulting bonds, in general, have the same credit rating the company. Yet, a company cab also be interpreted as the receiver of future cash flows with different credit cahs flows, NOt all the receivables will have the same rating. For example, some cash flows may be owed by institutions with dubious credit record, and these cash flows may not be received in the case of default or delinquency. Other cash flows may be liabilities of highly reputable companies, may carry a law probability of default , and may indeed be receive with very high probability.

Yet, a debt issue will be backed by an average of these credits, since it is the average receivable cash flows that determine the probability that the bonds will be repaid at maturity. If the receivables of a company carry mostly a relatively high probability of default, then the company may experience difficulties in the future and, hence, may end up defaulting onthe loan. Alternatively, the credit spread on the bond will increase and the investor will be subject to mark-to-market losses. All these possibilities reflect on the debt issued by this company, and, are factor in the determination of the proper cost of funding.

On the othe hand, instead of issuing debt on the back of the average cash flows to be received in the future, the company cab issue special types of bonds that are backed only be the higher-rated portion of the receivables. Clearly , such receivables have a comparatively lower probability of default, and this makes the bonds carry a lower default probability. The funding cost will decrease significantly. The Company has thus securitized a certain portion of the cash flows that are to be received in the future. In other world, securitization cab be regarded as a way to issue debt and raise funds that have a higher rating than that of the company. It is also a way of repackaging various cash flows.

What are the critical aspects of such financial engineering? Essentially, various cash flows are to be analyzed and a proper selection is made so as to obtain an optimal basket. This is then sold to investors through types of bonds.

Yet, besides the financial engineering aspects, securitization involves (1) legal issues, (2) balance sheet considerations, and (3) tax considerations. Securitization is a way of funding an operation. Instead of selling bonds or securing bank lines, the company issues asset-backed securities. The option of securitization helps corporate and banks to make decisions among the various funding alternatives.

4.1 Choosing Cash Flows

Consider Figure 17-7. Here, we show a bank that expects three different (random) cash flows in the future. The institutions that are suppose to pay these cash flows have different credit ratings. For example, the first series of cash flows, rated BBB, may represent credit card payments. The third could represent the random cash flow due to mortgages than they do to the timely payments of unsecured credit card proceeds. Credit card defaults are much common and plausible than mortgage than they do to the timely payments of unsecured credit card proceeds. Credit card defaults are much more common and plausible than mortgage delinquencies. Thus, the mortgage cash flows will be rated higher, say, with an A rating as shown in the figure.

Now, if the company is set to receive these three cash flows only, assuming similar liabilities, the company's average rating will perhaps be around BBB+ a Corporate bond issued by the company will carry a BBB+ credit spread.

Consider two different ways of packaging the same cash flow. If the company "sells" cash flow 3 and backs a bond issue with this cash flow only, the probability of default will be much lower and funding cab be secured at a lower rate. A bond backed by cash flows 1 and 3 will have a lower credit rating but still yields a funding cost below that of a general bond issue. The funding cos twould be a little higher, but at the same time , more cab be borrowed because ther is a bigger pool of receivables in this second option.

4.2 The Critical Stop: Securing the Cash Flow

The idea of securitization is quite simple. Instead of borrowing agains the average quality of the company's receivables, which is what happens if the company sells a straight corporate bond, the entry decides to borrow against a higher quality subset of the receivables. In the case of default , these receivables hav a higher change of being collected (recovered) and, hence, the cost of these funds will be lower.

But there is a critical step. How cab the buyer of an asset-backed security make sure that the receivables that are suppose to back the security are not used by the company for other purposes, and that, in the case of bankruptcy, these receivables will be there to cover losses?

The question is relevant, since after issuing the ABS security, it is still the original company that handles the business of processing new receviables (e.g. , by issuing new mortgages) , as well as the receipts of cash generated by such cashflows, and then uses them in the daily business of the firm. Clearly , there must be an additional mechanism that guarantees , at least partially, that these cash flows will be there in case of default.

A bankruptcy remote SPV is one such mechanism used quite often to resolve such problems (SPV), which is a shell company, often independent of the parent company, and whose sole purpose is to act as a vehicle in structuring the ABS (2) Steps are taken to make the SPV bankruptcy remote. That is to say, the probability that the SPV itself defaults is zero (since it does not engage in any meaningful economic activity other than that of issuing the paper), and in case the original company goes bankrupt, the underlying cash flows remain in the hand of SPV (3) the issuing company draws all the necessary papers so that, at least from a legal point of view, the cash flows are sold to to SPV. This is a true sale at law. [12]

The idea is to transfer the right to these cash flows and guarantee them under the ABS as much as possible. In fact, several SPVs with different purposes cab be layered to make sure that the ABS has the desired characteristics.

1. An SPV may be needed for tax resons.
2. Another SPV may needed for balance sheet reason.
3. Still another SPV may be needed to comply with other regulations.

Hence, one possible structure can be layered as shown in Figure 17-8 . Note that , here , the first SPV is a subsidiary of the company , so the company cab "buy" the cash flows, and this is the reason for its existence , But if the SPVS keeps the cash flows, these will still be on the balance sheet of the original company.

Finally, note the role played by the investment bank. The first three layers make up the ABS structure , and the investment bank still has to handle the original sale of the ABS. The structure clearly shows that the ABS has three important purposes, namely lowering the funding cost, managing the balance sheet, and handling tax and accounting restrictions.

4.1 Some Comparisons

The first use of securitization concerns funding costs, as already discussed. Securitization is a from of funding. But we must add that it is also an unconstrained form of funding, and an off-balance sheet form of liquidity for small and medium companies,. Finally, it is a diversified funding source. This way it can lower leverage. Securitization also implies less public disclosure.

Securitization is neither secured corporate financing , not the sale of an aset, It is hybrid, a combination of both that uses the will accepted legal, regulatory, tax, and accounting concepts that already exist.

4.3.1 Loan –sales

We should also compare whole-loan sales versus securitization. Securitization is on a service-retained basis, where as loan sales is service released. The buyer of the loan would like to service the loan himself or herself.

A second point is the retention of credit and prepaid risk. In a loan sale, 100% of these risk are transferred. With securitization, some of these risk may be retained.

Third, a loan sale is often done at a premium, whereas securitization issues are often around par.

Fourth, there is a timing issue. In securitization, cash flow from assets are often invested in short-term investments are then transferred to the bondholders. Thus, the investor receives the payments late than the servicer. Finally, securitization sometimes uses credit-enhancements and this makes the paper somewhat more liquid.

4.3.1 Secured Lending

Securitization is similar to secured financing, with one important difference. In an abs, the issuing company is not liable for its assets backed securities. It is as if the company has not really "borrowed" the funds. A separate legal entity needs to be established to do the borrowing. Securitization is structured so that this entity becomes the legal owner of the asset. If the company defaults, the cash flows will belong to the company, but to the SPV. This way the owners of the bonds have an ownership in the case of securitization, whereas in the case of secured lending, they only have a security interest.

DISCUSSIONS

EIILM UNIVERSITY
SIKKIM

Jorethang, District Namchi, Sikkim- 737121, India
www.eiilmuniversity.ac.in