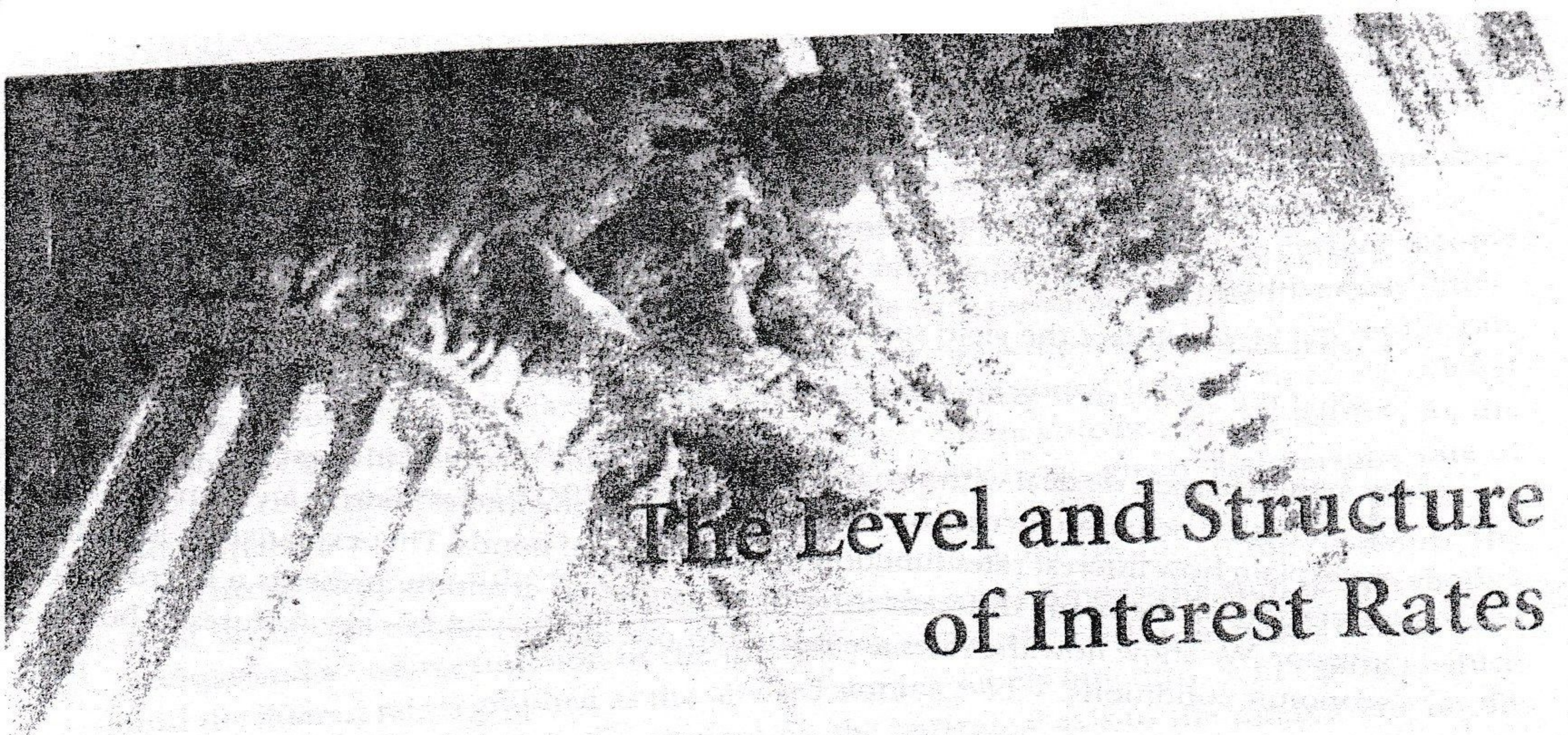


The Level and Structure of Interest Rates

LEARNING OBJECTIVES

After reading this chapter, you will understand

- Fisher's classical approach to explaining the level of the interest rate
- the role in Fisher's theory of the saver's time preference and the borrowing firm's productivity of capital
- the meaning of equilibrium and how changes in the demand and supply function affect the equilibrium level of the interest rate
- the structure of Fisher's Law, which states that the nominal and observable interest rate is composed of two unobservable variables: the real rate of interest and the premium for expected inflation
- the loanable funds theory, which is an expansion of Fisher's theory
- the meaning of liquidity preference in Keynes's theory of the determination of interest rates
- how an increase in the money supply can affect the level of the interest rate through an impact on liquidity, income, and price expectations
- the features of a bond issue
- how the yield to maturity of a bond is calculated
- why historically the yields on securities issued by the U.S. Department of the Treasury have been used as the benchmark interest rates throughout the world
- the reason why the yields on U.S. Treasury securities are no longer as popular as benchmark interest rates and what alternative benchmarks are being considered by market participants

From Chapter 10 of *Foundations of Financial Markets and Institutions*, 4/e. Frank J. Fabozzi. Franco Modigliani. Frank J. Jones.
Copyright © 2010 by Pearson Prentice Hall. All rights reserved.

- the different types of bonds
- what factors affect the yield spread between two bonds
- what the swap curve is and why it is used as an interest rate benchmark

In this chapter, we deal with two issues. The first is the theory of interest rate determination, and our discussion focuses on the general level of the interest rate in an economy. Second, we explain how interest rates function in the pricing of bonds. That explanation leads to consideration of the many different interest rates that an economy presents a borrower or an investor. We show how the rates are related to one another and to key features of bonds and economic conditions.

THE THEORY OF INTEREST RATES

An interest rate is the price paid by a borrower (or debtor) to a lender (or creditor) for the use of resources during some interval. The amount of the loan is the *principal*, and the price paid is typically expressed as a percentage of the principal per unit of time (generally, a year). In this section, we present the two most influential theories of the determination of the interest rate: **Fisher's theory of interest**, which underlies the loanable funds theory, and Keynes's **liquidity preference theory of interest**.

We focus first on the interest rate that provides the anchor for other rates, namely, the short-term, risk-free, real rate. By the *real rate*, we mean the rate that would prevail in the economy if the average prices for goods and services were expected to remain constant during the loan's life. By the *risk-free rate*, we mean the rate on a loan whose borrower will not default on any obligation. By *short-term*, we mean the rate on a loan that has one year to maturity. All other interest rates differ from this rate according to particular aspects of the loan, such as its maturity or risk of default, or because of the presence of inflation.

Fisher's Classical Approach

Irving Fisher analyzed the determination of the level of the interest rate in an economy by inquiring why people save (that is, why they do not consume all their resources) and why others borrow.[1] Here, we outline his theory in the context of a very simplified economy. That economy contains only individuals who consume and save with their current income, firms that borrow unconsumed income in loans and invest, a market where savers make loans of resources to borrowers, and projects in which firms invest. The interest rate on loans embodies no premium for default risk because borrowing firms are assumed to meet all obligations. (The prospect of inflation and its impact on the interest rate will be discussed in short order.)

Decisions on Saving and Borrowing

Saving is the choice between current and future consumption of goods and services. Individuals save some of their current income in order to be able to consume more in the future. A chief influence on the saving decision is the individual's **marginal rate of time preference**, which is the willingness to trade some consumption now for more future consumption. Individuals differ in their time preferences. Some people may have a rate of time preference

[1] Irving Fisher, *The Theory of Interest Rates* (New York: Macmillan, 1930).

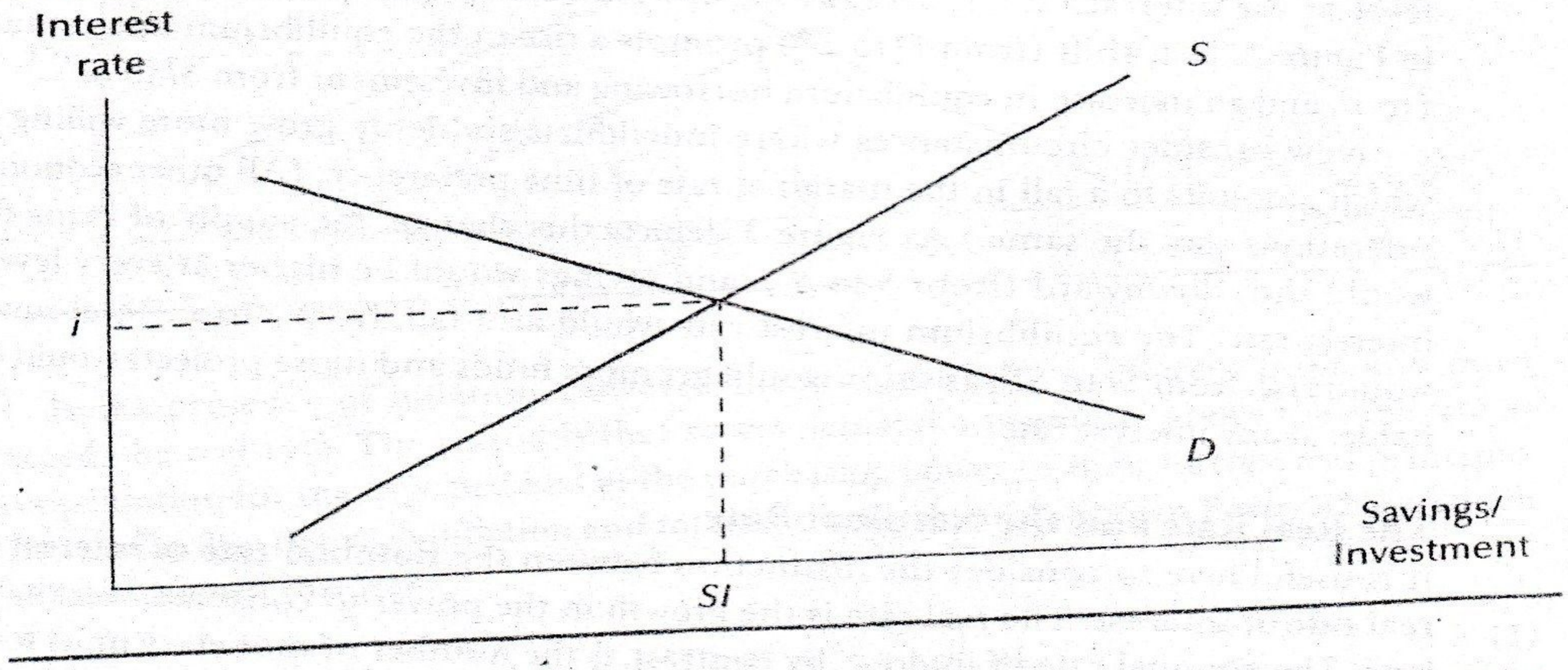
that leads them to forgo current consumption for an increase of 10% in their future consumption, while others might save only if their future consumption possibilities rise by 20%.

Another influence on the saving decision is *income*. Generally, higher current income means the person will save more, although people with the same income may have different time preferences. The third variable affecting savings is the *reward for saving*, or the rate of interest on loans that savers make with their unconsumed income. Interest is what borrowers pay for the loans, and it makes greater future consumption possible. As the interest rate rises, each person becomes willing to save more, given that person's rate of time preference.

This description of the savings decision applies to all the people in the economy. The total savings (or the total supply of loans) available at any time is the sum of everybody's savings and a positive function of the interest rate. The relationship between total savings and the interest rate is graphed as the upward sloping supply function, *S*, in Figure 1, which relates the amount of savings/investment on the horizontal axis to the interest rate on the vertical axis.[2]

There can be no reward for savings if there is no demand for borrowed resources because someone must pay the interest. In our simple economy, firms do all the borrowing, and they borrow from savers in order to invest. Investment means directing resources to assets that will increase the firms' future capacity to produce. An important influence on the borrowing decision is the *gain from investment*, which is the positive difference between the resources used by a process and the total resources it will produce in the future. The gain is not constant, at any one time, across all possible projects and levels of total investment. The reason is that, at any time, only a certain number of projects are available, some offering high gains, others promising moderate gains, and still others yielding low gains. Firms will direct borrowed resources to projects in order of their profitability, starting with

Figure 1 Equilibrium in the Market for Savings



[2] For a more extensive analysis of the supply of savings, including the reason why the supply can reach negative levels and the reason why the supply curve can itself turn negative, see Frank J. Fabozzi and Franco Modigliani, *Capital Markets: Institutions and Instruments* (Englewood Cliffs, N.J.: Prentice Hall, 1992), pp. 338ff.

the most profitable and proceeding to those with lower gains. The gain from additional projects, as investment increases, is the **marginal productivity of capital**, which is negatively related to the amount of investment. In other words, as the amount of investment grows, additional gains necessarily fall, as more of the less profitable projects are accepted.

The maximum that a firm will invest depends on the *rate of interest*, which is the cost of loans. The firm will invest only as long as the marginal productivity of capital exceeds or equals the rate of interest. In other words, firms will accept only projects whose gain is not less than their cost of financing. Thus, the firm's demand for borrowing is negatively related to the interest rate. If the rate is high, only limited borrowing and investment make sense. At a low rate, more projects offer a profit, and the firm wants to borrow more. This negative relationship exists for each and all firms in the economy. The economy's total demand for borrowed resources (or loans of unconsumed income), as a function of the interest rate, appears as the downward-sloping line labeled D in Figure 1.

Equilibrium in the Market

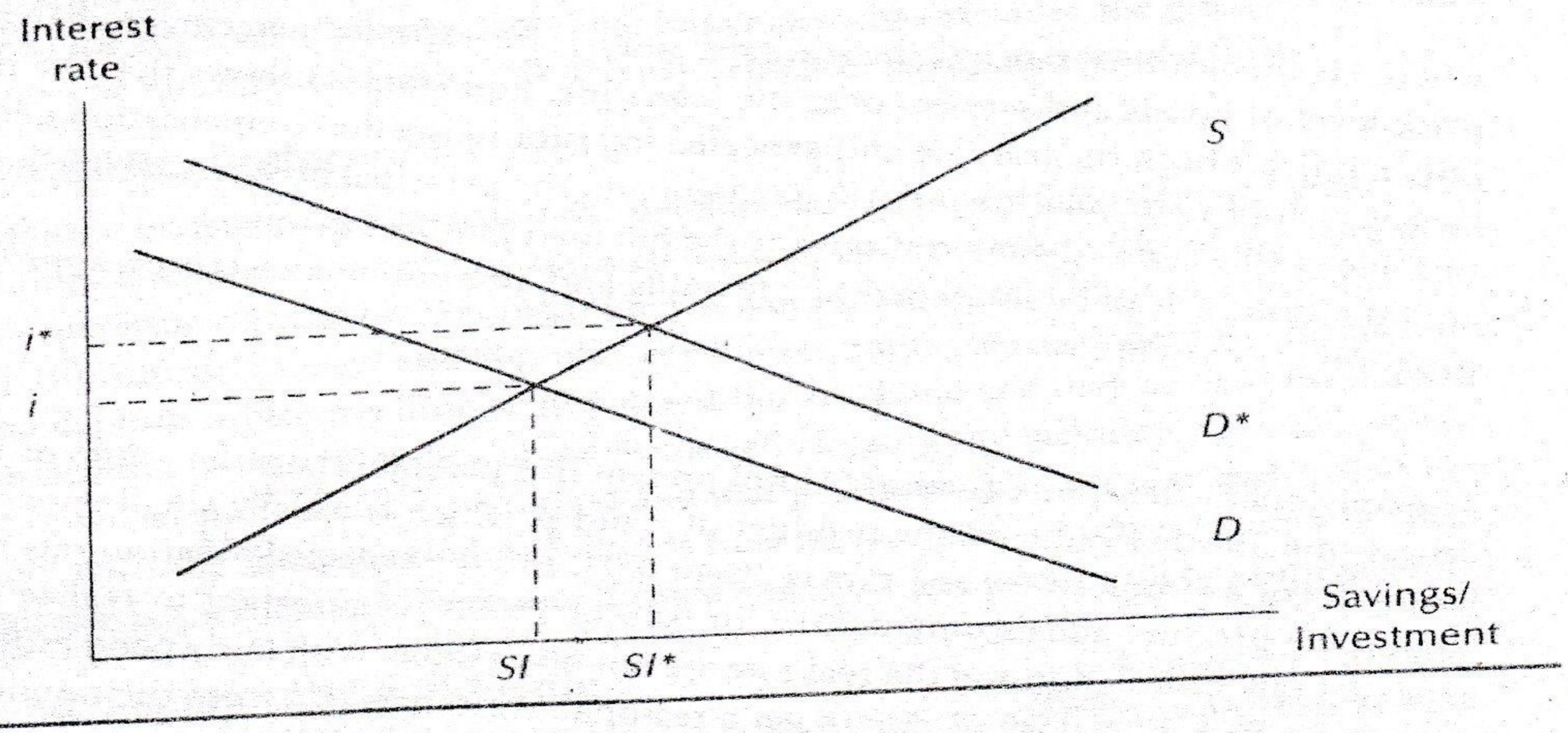
The equilibrium rate of interest is determined by the interaction of the supply and demand functions. As a cost of borrowing and a reward for lending, the rate must reach the point where total supply of savings equals total demand for borrowing and investment. Figure 1 shows that this equilibrium rate of interest, labeled i, occurs at the intersection of the demand and supply curves, D and S. The equilibrium level of savings (which is the same as the equilibrium level of borrowing and investment) is given as SI. Clearly, Fisher's theory emphasizes that the long-run level of the interest rate and the amount of investment depend on a society's propensity to save and on technological development.

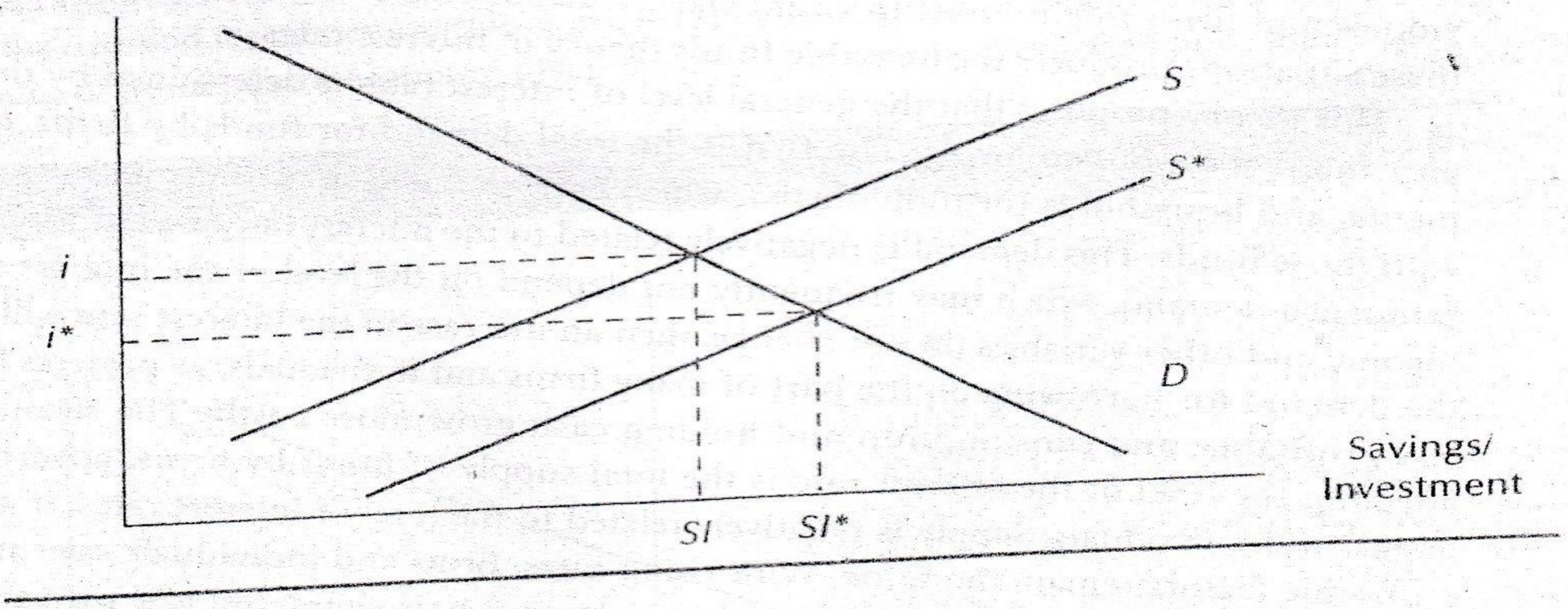
Let us now consider the effects of a sudden increase in technological capability, which makes production cheaper. With no change in any other relevant variable, lower production costs mean more gain on investments and a higher marginal productivity of capital. The resulting increase in firms' desired investment and borrowing through loans, at any level of the interest rate, is actually an upward shift in the demand function, as shown in Figure 2. That shift (from D to D^*) prompts a rise in the equilibrium interest rate, from i to i^*, and an increase in equilibrium borrowing and investment from SI to SI^*.

Now consider circumstances where individuals suddenly grow more willing to save, which amounts to a fall in the marginal rate of time preference. (All other economic considerations stay the same.) As Figure 3 depicts this change, the supply of loans function would shift downward (from S to S^*), and savings would be higher at every level of the interest rate. The equilibrium interest rate would also fall, from i to i^*. Total investment would rise, from SI to SI^*, as firms would get more funds and more projects would be profitable, at any interest rate.

The Real Rate and the Nominal Rate

It is useful here to consider the distinction between the **nominal rate of interest** and the **real rate of interest.** The real rate is the growth in the power to consume over the life of a loan. The nominal rate of interest, by contrast, is the number of monetary units to be paid per unit borrowed, and is, in fact, the observable market rate on a loan. In the absence of inflation, the nominal rate equals the real rate. Because we assume in our discussion that the purchasing power of the monetary units in a loan stays the same over the loan's life, the equilibrium interest rate we have discussed is both the nominal and the real rate of interest.





In the presence of inflation, however, the nominal rate is different from, and must exceed, the real rate. The reason is that savers demand a premium above the real rate as compensation for the expected loss in the purchasing power of their interest and principal. The relationship between inflation and interest rates is the well-known **Fisher's Law**, which can be expressed this way:

$$(1 + i) = (1 + r) \times (1 + p), \tag{1}$$

where i is the nominal rate, r is the real rate, and p is the expected percentage change in the price level of goods and services over the loan's life. Equation (1) shows that the nominal rate, i, reflects both the real rate and expected inflation. Also, the compensation demanded by savers applies to both the interest payment and the principal because the left-hand side variable is 100% plus the interest rate, or the full loan plus its interest.

Equation (1) can be simplified to a formula and, in most circumstances, approximates equation (1) closely enough:

$$i = r + p \tag{2}$$

The only quantities from equation (1) not present in equation (2) are the values of 1, which cancel one another out, and the product of r and p, which is usually small enough to be ignored. For example, if the real rate is 3% or 0.03, and the expected inflation rate is 5% or 0.05, their product equals only 0.0015, or 0.15 of 1%. It is important to realize that the expected rate of inflation and the real rate are not observable. With fairly good estimates of expected inflation, p, it is possible to get a reasonable estimate of r from the nominal rate, but the precise value of the real rate must remain elusive.

The Loanable Funds Theory

Fisher's theory is a general one and obviously neglects certain practical matters, such as the power of the government (in concert with depository institutions) to create money and the government's often large demand for borrowed funds, which is frequently immune to the level of the interest rate. Also, Fisher's theory does not consider the possibility that individuals and firms might invest in cash balances. Expanding Fisher's theory to encompass these situations produces the **loanable funds theory of interest rates.**

This theory proposes that the general level of interest rates is determined by the complex interaction of two forces. The first is the total demand for funds by firms, governments, and households (or individuals), which carry out a variety of economic activities with those funds. This demand is negatively related to the interest rate (except for the government's demand, which may frequently not depend on the level of the interest rate). If income and other variables do not change, then an increase in the interest rate will reduce the demand for borrowing on the part of many firms and individuals, as projects become less profitable, and consumption and holding cash grow more costly. The second force affecting the level of the interest rate is the total supply of funds by firms, governments, banks, and individuals. Supply is positively related to the level of interest rates, if all other economic factors remain the same. With rising rates, firms and individuals save and lend more, and banks are more eager to extend more loans. (A rising interest rate probably does not significantly affect the government's supply of savings.)

In an equilibrium situation much like that depicted in Figure 1, the intersection of the supply and demand functions sets the interest rate level and the level of loans. In equilibrium, the demand for funds equals the supply of funds. This means that all agents are borrowing what they want, investing to the desired extent, and holding all the money they wish to hold. In other words, equilibrium extends through the money market, the bond market, and the market for investment assets.

As in Fisher's theory, shifts in the demand and supply curves may occur for many reasons: changes in the money supply, government deficits, changed preferences by individuals, new investment opportunities, and so on. These shifts affect the equilibrium level of the interest

rate and of investment in predictable ways. Finally, the expectation of inflation can affect the equilibrium rate through the supply of funds curve, as savers demand higher rates (because of inflation) for any level of savings. Note that this analysis has excluded the question of default on loans: The rate discussed is the risk-free rate, either in its real or nominal form.

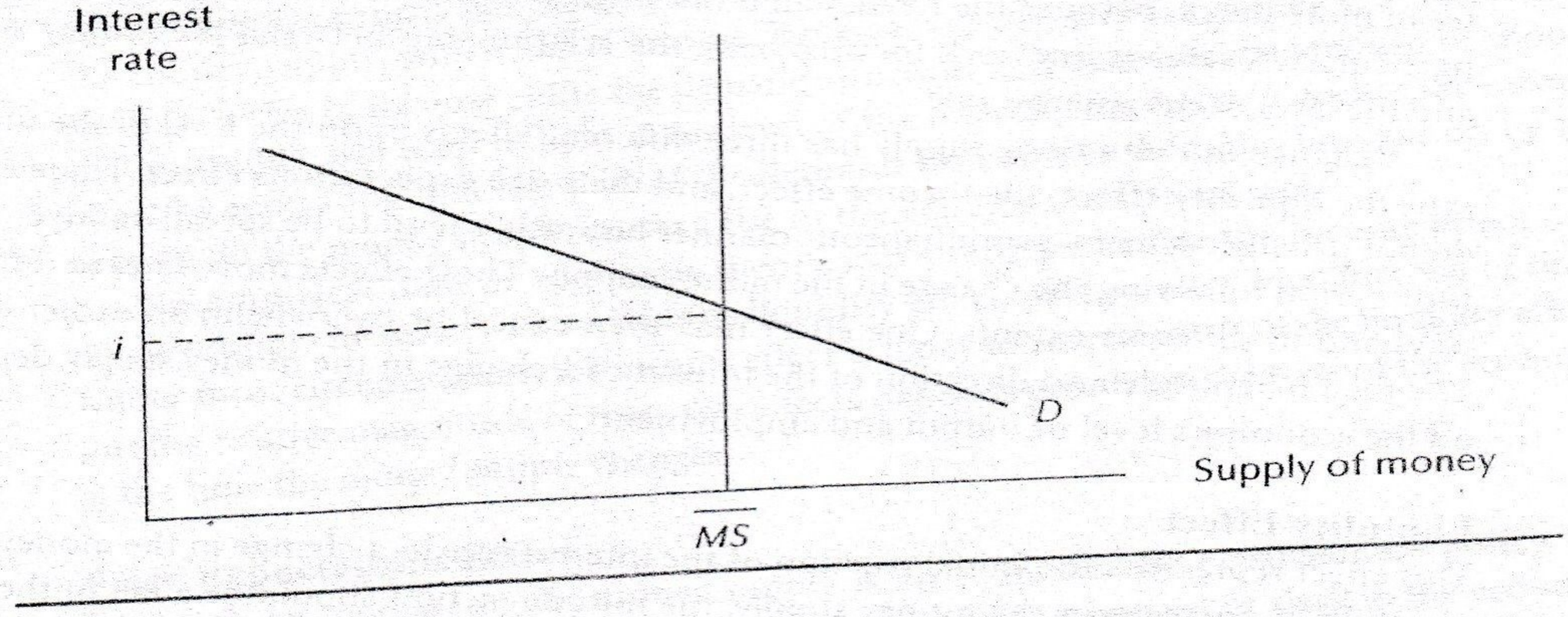
The Liquidity Preference Theory

The liquidity preference theory, originally developed by John Maynard Keynes,[3] analyzes the equilibrium level of the interest rate through the interaction of the supply of money and the public's aggregate demand for holding money. Keynes assumed that most people hold wealth in only two forms: "money" and "bonds." For Keynes, money is equivalent to currency and demand deposits, which pay little or no interest but are liquid and may be used for immediate transactions. Bonds represent a broad Keynesian category and include long-term, interest-paying financial assets that are not liquid and that pose some risk because their prices vary inversely with the interest rate level. Bonds may be liabilities of governments or firms. (Default risk is not considered here, and the rate is the risk-free rate, in real or nominal form.)

Demand, Supply, and Equilibrium

The public (consisting of individuals and firms) holds money for several reasons: ease of transactions, precaution against unexpected events, and speculation about possible rises in the interest rate. Although money pays no interest, the demand for money is a negative function of the interest rate. At a low rate, people hold a lot of money because they do not lose much interest by doing so and because the risk of a rise in rates (and a fall in the value of bonds) may be large. With a high interest rate, people desire to hold bonds rather than money, because the cost of liquidity is substantial in terms of lost interest payments and because a decline in the interest rate would lead to gains in the bonds' values. The negative linkage between the interest rate and the demand for money appears as curve D in Figure 4, which relates the interest rate to the amount of money in the economy, given the level of income and expected price inflation.

Figure 4 Equilibrium in Keynes's Market for Money



[3] John Maynard Keynes, *The General Theory of Employment, Interest and Money* (New York: Harcourt, Brace & World, 1936).

For Keynes, the supply of money is fully under the control of the central bank (which is the Fed in the United States). Moreover, the money supply is not affected by the level of the interest rate. (There is a positive link between the interest rate and the growth in the money supply.) Thus, the supply of money appears, in Figure 4, as the vertical line, $\overline{MS}$, and the line above the MS indicates a quantity not varying with the interest rate. Equilibrium in the money market requires, of course, that the total demand for money equals total supply. In Figure 4, equilibrium implies an interest rate of i. Furthermore, equilibrium in the money market implies the equilibrium of the bond market.

Shifts in the Rate of Interest

The equilibrium rate of interest can change if there is a change in any variable affecting the demand or supply curves. On the demand side, Keynes recognized the importance of two such variables: the level of income and the level of prices for goods and services. A rise in income (with no other variable changing) raises the value of money's liquidity and shifts the demand curve to the right, increasing the equilibrium interest rate. Because people want to hold amounts of "real money," or monetary units of specific purchasing power, a change in expected inflation would also shift the demand curve to the right and raise the level of interest.

The money supply curve can shift, in Keynes's view, only by actions of the central bank. The central bank's power over interest rates arises because of its ability to buy and sell securities (open market operations), which can alter the amount of money available in the economy. Generally, Keynes thought that an increase in the money supply would, by shifting the supply curve to the right, bring about a decline in the equilibrium interest rate. Similarly, he reasoned that a reduction in the money supply would raise rates. However, there is now widespread recognition that the question of linkage between the money supply and the level of the interest rate is rather more complex than that.

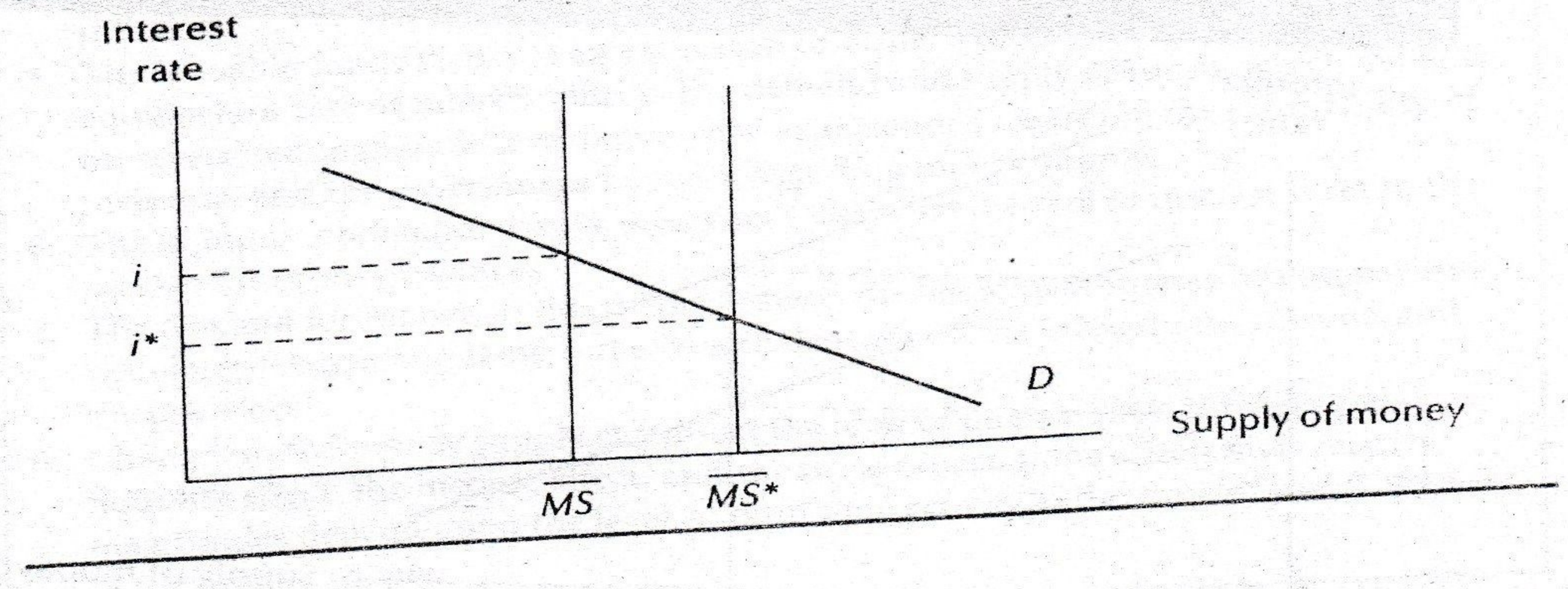
Changes in the Money Supply and Interest Rates

Changes in the money supply affect the level of interest rates. Now, we can discuss the matter in more detail, because the Keynesian money demand model that we just developed is a particularly useful framework for analyzing the relationship between the money supply and the level of the interest rate.

A change in the money supply has three different effects upon the level of the interest rate: the **liquidity effect**, the **income effect**, and the **price expectations effect**. These effects do not usually occur in a simultaneous manner but rather tend to be spread out over some time period following the change in the money supply. These effects move rates in different ways and to different extents. One effect may even cancel or overwhelm an earlier effect. The final magnitude and direction of the impact of a change in the money supply depends on the economy's level of output and employment.

Liquidity Effect

This effect represents the initial reaction of the interest rate to a change in the money supply. With an increase in the money supply, the initial reaction should be a fall in the rate. The reason for the fall is that a rise in the money supply represents a shift in the supply



curve. For example, let us suppose that (in the United States) the Fed increases the supply of money by buying bonds, raising excess reserves, and allowing banks to offer more loans. With demand unchanging, the increase in the money supply amounts to a rightward shift of the supply and causes a fall in the equilibrium interest rate, from i to i^*. Figure 5 depicts the rise in the money supply (from $\overline{MS}$ to $\overline{MS^*}$) and the decline in the interest rate from i to i^*. (A decrease in the money supply would cause the supply of money function to shift to the left and the liquidity effect of an initial rise in the interest rate.)

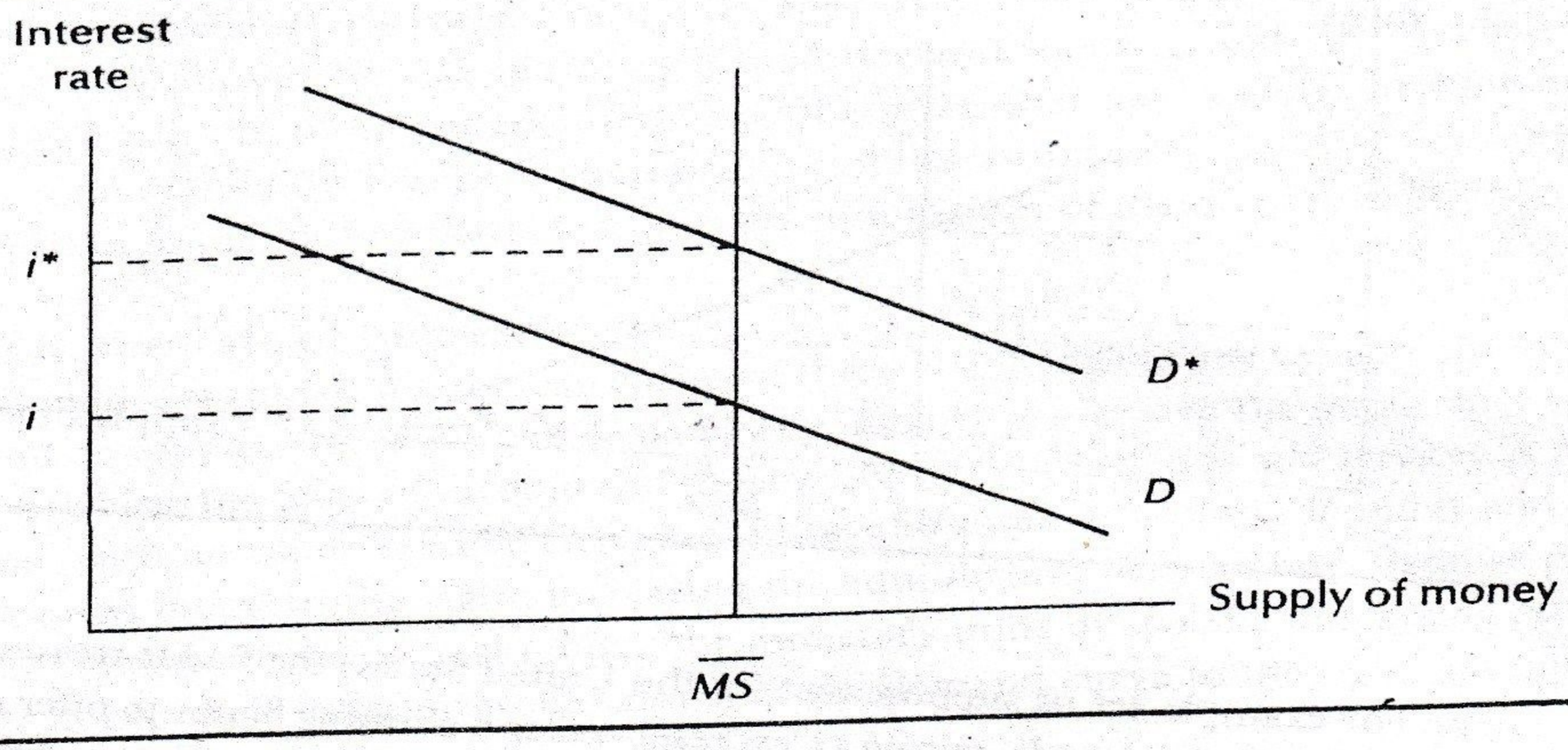
Income Effect

It is well known that changes in the money supply affect the economy. A decline in the supply would tend to cause a contraction. An increase in the money supply, generally speaking, is economically expansionary: More loans are available and extended, more people are hired or work longer, and consumers and producers purchase more goods and services. Thus, money supply changes can cause income in the system to vary. Let us focus on an increase in the money supply, which raises income. Because the demand for money is a function of income, a rise in income shifts the demand function and increases the amount of money that the public will want to hold at any level of the interest rate. Figure 6 depicts the income effect by indicating that the shift of the demand function to the right, brought on by the increase in the money supply, causes a rise in the equilibrium interest rate.

No empirical data or economic theory can predict whether the income effect of a money supply increase will override its liquidity effect, or if so, after what interval of time. Under most circumstances, the income effect is likely to reverse some of the liquidity effect. But the relative magnitude of these two effects depends greatly on the state of the economy at the time the money supply changes.

Price Expectations Effect

Although an increase in the money supply is an economically expansionary policy, the resultant increase in income depends substantially on the amount of slack in the economy at the time of the Fed's action. If the economy is operating at less than full strength, an increase in the money supply can stimulate production, employment, and output; if the



economy is producing all or almost all of the goods and services it can (given the size of the population and the amount of capital goods), then an increase in the money supply will largely stimulate expectations of a rising level of prices for goods and services. Thus, the price expectations effect usually occurs only if the money supply grows in a time of high output.

Because the price level (and expectations regarding its changes) affects the money demand function, the price expectations effect is an increase in the interest rate. That increase occurs because the demand for money balances shifts upward. This positive effect moves the interest rate in the same direction as the income effect, and in the opposite direction of the liquidity effect. (Of course, in a time of inflation, a reduction in the rate of growth in the money supply might dampen inflationary expectations and shift the demand curve to the left, leading to declines in the interest rate.)

There is no general guide to the relative size of the price expectations effect: It may be great enough to overwhelm the liquidity effect, or it may cancel only part of it. The magnitude of the income effect depends upon how much of the economy's productive capacity it is utilizing when the money supply rises.

Key Points That You Should Understand Before Proceeding

1. Interest is the price paid for the temporary use of resources, and the amount of a loan is its principal.
2. Fisher's theory of interest analyzes the equilibrium level of the interest rate as the result of the interaction of savers' willingness to save and borrowers' demand for investment funds.
3. In Fisher's terms, the interest rate reflects the interaction of the savers' marginal rate of time preference and borrowers' marginal productivity of capital.

4. Fisher's Law states that the observable nominal rate of interest is composed of two unobservable variables: the real rate of interest and the premium for expected inflation.
5. The loanable funds theory is an extension of Fisher's theory and proposes that the equilibrium rate of interest reflects the demand and supply of funds, which depend on savers' willingness to save, borrowers' expectations regarding the profitability of investing, and the government's action regarding money supply.
6. The liquidity preference theory is Keynes's view that the rate of interest is set in the market for money balances.
7. The demand for money, in this theory, reflects the liquidity of money by comparison with long-term financial instruments and depends on the interest rate, income, and the price level.
8. Changes in the money supply can affect the level of interest rates through the liquidity effect, the income effect, and the price expectations effect; their relative magnitudes depend upon the level of economic activity at the time of the change in the money supply.

THE DETERMINANTS OF THE STRUCTURE OF INTEREST RATES

There is not one interest rate in any economy. Rather, there is a structure of interest rates. The interest rate that a borrower will have to pay will depend on a myriad of factors. Although we describe these various factors in the process, we provide an overview of the fixed-income instruments. We begin with a basic description of a bond.

Features of a Bond

A bond is an instrument in which the issuer (debtor/borrower) promises to repay to the lender/investor the amount borrowed plus interest over some specified period of time. The **term to maturity** of a bond is the number of years during which the issuer has promised to meet the conditions of the obligation. The maturity of a bond refers to the day the debt will cease to exist and the day the issuer will redeem the bond by paying the amount owed. The practice in the bond market, however, is to refer to the *term to maturity* of a bond as simply its *maturity* or *term*. As we explain in the following text, there may be provisions that allow either the issuer or bondholder to alter a bond's term to maturity.

The **principal value** (or simply **principal**) of a bond is the amount that the issuer agrees to repay the bondholder at the maturity date. This amount is also referred to as the par **value, maturity value, redemption value,** or **face value.**

The *coupon rate* is the interest rate that the issuer agrees to pay each year. The annual amount of the interest payment made to owners during the term of the bond is called the *coupon*. The coupon rate when multiplied by the principal of the bond provides the dollar amount of the coupon. For example, a bond with an 8% coupon rate and a principal of $1,000 will pay annual interest of $80. In the United States and Japan, the usual practice is for the issuer to pay the coupon in two equal semiannual installments. In contrast, bonds issued in most European bond markets and the Eurobond market make coupon payments only once per year.

Yield on a Bond

The yield on a bond investment should reflect the coupon interest that will be earned plus either (1) any capital gain that will be realized from holding the bond to maturity, or (2) any capital loss that will be realized from holding the bond to maturity. For example, if a four-year bond with a coupon rate of 5% and a par value of $1,000 is selling for $900.64, the yield should reflect the coupon interest of $50 (5% times $1,000) every year plus the capital gain of $99.36 ($1,000 – $900.64) when the bond is redeemed at maturity. Moreover, the yield should also reflect the time value of money by considering the timing of the various payments associated with the bond.

The **yield to maturity** is a formal, widely accepted measure of the rate of return on a bond. As typically defined, the yield to maturity of a bond takes into account the coupon interest and any capital gain or loss, if the bond were to be held to maturity. The yield to maturity is defined as the interest rate that makes the present value of the cash flow of a bond equal to the bond's market price. In mathematical notation, the yield to maturity, y, is found by solving the following equation for y:

$$P = \frac{C}{(1+y)^1} + \frac{C}{(1+y)^2} + \frac{C}{(1+y)^3} + \cdots + \frac{C+M}{(1+y)^n}$$

where

P = market price of bond
C = coupon interest
M = maturity value
n = time to maturity

The yield to maturity is determined by a trial-and-error process. Even the algorithm in a calculator or computer program, which computes the yield to maturity (or internal rate of return) in an apparently direct way, uses a trial-and-error process. The steps in that process are as follows:

Step 1: Select an interest rate.
Step 2: Compute the present value of each cash flow using the interest rate selected in Step 1.
Step 3: Total the present value of the cash flows found in Step 2.
Step 4: Compare the total present value found in Step 3 with the market price of the bond and, if the total present value of the cash flows found in Step 3 is

- equal to the market price, then the interest rate used in Step 1 is the yield to maturity
- greater than the market price, then the interest rate is not the yield to maturity. Therefore, go back to Step 1 and use a higher interest rate
- less than the market price, then the interest rate is not the yield to maturity. Therefore, go back to Step 1 and use a lower interest rate

For example, suppose that a four-year bond that pays interest annually has a par value of $1,000, a coupon rate of 5%, and is selling in the market for $900.64. Table 1 shows the procedure for calculating the yield to maturity. When a 2% interest rate is tried, the total present value is equal to $1,136.52. As this value is greater than the market price of $900.64, it is not the yield to maturity and, therefore, a higher interest rate must be tried.

$C = \$50$ $M = \$1{,}000$ $P = \$900.64$ $n = 4$

Interest Rate Tried	Total Present Value
2%	$1,136.52
3	1,106.57
4	1,036.30
5	1,000.00
6	965.35
7	932.26
8	900.64
9	870.41
10	841.51
11	813.85
12	787.39

Suppose a 12% interest rate is tried. Now the total present value of $787.39 is smaller than the market price, so a lower interest rate must be tried. When an 8% interest rate is tried, the total present value is equal to the market price, so the yield to maturity is 8%.

Had the market price of this bond been $813.85 rather than $900.64, the yield to maturity would be 11%. Notice that the lower the market price, the higher the yield to maturity.

There are several interesting points about the relationship among the coupon rate, market price, and yield to maturity that can be seen in Table 1.

1. If the market price is equal to the par value, then the yield to maturity is equal to the coupon rate.
2. If the market price is less than the par value, then the yield to maturity is greater than the coupon rate.
3. If the market price is greater than the par value, then the yield to maturity is less than the coupon rate.

In our illustration, we assumed that coupon payments are made once per year. As noted earlier, however, the practice in the U.S. bond market is to pay coupon interest every six months. This does not affect the procedure for calculating the yield to maturity; that is, it is still the interest rate that will make the present value of the cash flow equal to the market price. The cash flow, however, is not annual. Instead, the coupon payment is every six months and is equal to one-half the annual coupon payment. The resulting yield to maturity is then a semiannual yield to maturity. To annualize the semiannual yield, the *convention* adopted in the bond market is to double the semiannual yield. The resulting yield to maturity is said to be calculated on a **bond-equivalent yield basis.**

For example, consider an 18-year bond with a coupon rate of 6% and a par value of $1,000. Suppose that the bond pays interest semiannually and is selling for $700.89. The cash flow for this bond is $30 every six months for 35 six-month periods and $1,030 36 six-month periods from now. That is,

$C = \$30$ $M = \$1{,}000$ and $n = 36$ (double the number of years to maturity)

Table 2 shows the calculation of the yield to maturity for this bond. Notice that the interest rate that makes the total present value equal to the market price of $700.89 is 4.75%. This is the semiannual yield to maturity. Doubling this semiannual yield gives 9.5%, which is the yield to maturity on a bond-equivalent yield basis.

In the discussion to follow, we will refer to the yield to maturity of a bond as simply its yield. The difference between the yield on any two bond issues is called a yield spread, or simply **spread**. The spread is typically measured in basis points. For example, if the yield on bond A is 9%, and the yield on bond B is 8.5%, the yield spread is 50 basis points.

The Base Interest Rate

Now that we have examined the characteristics of a bond and how to measure the yield on a bond, we will look at the structure of interest rates. In our discussion, we use the terms *interest rate* and *yield* interchangeably.

The securities issued by the U.S. Department of the Treasury, popularly referred to as **Treasury securities**, or simply **Treasuries**, are backed by the full faith and credit of the U.S. government. Consequently, market participants throughout the world view them as having no credit risk. As a result, historically the interest rates on Treasury securities have served as the benchmark interest rates throughout the U.S. economy, as well as in international capital markets. Other important benchmarks have developed in international capital markets, and we discuss these later in this chapter.

Treasury securities are used to develop the benchmark interest rates. There are two categories of U.S. Treasury securities, discount and coupon securities. The fundamental difference between the two types is in the form of the stream of payments that the holder receives, which, in turn, is reflected in the prices at which the securities are issued. Coupon securities pay interest every six months, plus principal at maturity. Discount securities pay only a contractually fixed amount at maturity. Treasury securities are typically issued on an auction basis according to regular cycles for securities of specific maturities. Current Treasury practice is to issue all securities with maturities of one year or less as discount securities. These securities are called *Treasury bills*. All securities with maturities of two years or longer are issued as Treasury *coupon securities*.

$C = \$30$ $M = \$1{,}000$ $P = \$700.89$ $n = 36$

Interest Rate Tried	Total Present Value
3.25%	[illegible]
3.50	898.54
3.75	[illegible]
4.00	810.92
4.25	[illegible]
4.50	735.01
4.75	[illegible]

*Semiannual yield to maturity. Therefore, the yield to maturity on a bond-equivalent yield basis is 9.5% (2 times 4.75%).

The most recently auctioned Treasury issues for each maturity are referred to as **on-the-run Treasury issues** or **current coupon issues**. On May 20, 2008, the yield for four on-the-run issues as reported by the U.S. Department of the Treasury was

two-year Treasury 2.34%
five-year Treasury 3.02%
10-year Treasury 3.78%
30-year Treasury 4.53%

Issues auctioned prior to the current coupon issues are typically referred to as **off-the-run Treasury issues**; they are not as liquid as on-the-run issues, and, therefore, offer a higher yield than the corresponding on-the-run Treasury issue. Note that every day, the U.S. Department of the Treasury estimates the yield for Treasury securities based on the yield for on-the-run and off-the-run issues.[4]

The minimum interest rate or **base interest rate** that investors will demand for investing in a non-Treasury security is the yield offered on a comparable maturity on-the-run Treasury security. So, for example, if an investor wanted to purchase a 10-year bond on May 20, 2008, the minimum yield the investor would seek is 3.78%, the on-the-run Treasury yield reported above. The base interest rate is also referred to as the **benchmark interest rate.**

The Risk Premium

Market participants talk of interest rates on non-Treasury securities as "trading at a spread" to a particular on-the-run Treasury security (or a spread to any particular benchmark interest rate selected). For example, if the yield on a 10-year non-Treasury security on May 20, 2008, is 4.78%, then the spread is 100 basis points over the 3.78% Treasury yield. This spread reflects the additional risks the investor faces by acquiring a security that is not issued by the U.S. government and, therefore, can be called a **risk premium**. Thus, we can express the interest rate offered on a non-Treasury security as

$$\text{Base interest rate} + \text{Spread}$$

or equivalently,

$$\text{Base interest rate} + \text{Risk premium}$$

We have discussed the factors that affect the base interest rate. One of the factors is the expected rate of inflation. That is, the base interest rate can be expressed as

$$\text{Base interest rate} = \text{Real rate of interest} + \text{Expected rate of inflation}$$

How can the real rate of inflation required by market participants be estimated? Since 1997, the U.S. Department of the Treasury has issued securities indexed to the Consumer Price Index. These securities are called *Treasury Inflation Protection Securities* (TIPS).

Turning to the spread, the factors that affect it are (1) the type of issuer, (2) the issuer's perceived creditworthiness, (3) the term or maturity of the instrument, (4) provisions that grant either the issuer or the investor the option to do something, (5) the taxability of the interest received by investors, and (6) the expected liquidity of the issue. Notice that these factors are the ones that we discussed in the previous chapter as affecting the price of a security.

[4] http://www.treas.gov/offices/domestic-finance/debt-management/interest-rate/yield.shtml.

It is important to note that yield spreads must be interpreted relative to the benchmark interest rate used. This is particularly important to keep in mind for the second and last factors that affect the spread when the benchmark interest rate is other than the yield on U.S. Treasury securities.

Types of Issuers

A key feature of a debt obligation is the nature of the issuer. In addition to the U.S. government, there are agencies of the U.S. government, municipal governments, corporations (domestic and foreign), and foreign governments that issue bonds.

The bond market is classified by the type of issuer, and groups of securities of the various kinds of issuers are referred to as **market sectors**. The spread between the interest rates offered in two sectors of the bond market on obligations with the same maturity is referred to as an **intermarket sector spread.**

Excluding the Treasury market sector, the other market sectors include a wide range of issuers, each with different abilities to satisfy their contractual obligations. For example, within the corporate market sector, issuers are classified as follows: (1) utilities, (2) industrials, (3) finance, and (4) banks. The spread between two issues within a market sector is called an **intramarket sector spread.**

Perceived Creditworthiness of Issuer

Default risk or **credit risk** refers to the risk that the issuer of a bond may be unable to make timely principal or interest payments. Most market participants rely primarily on commercial rating companies to assess the default risk of an issuer. These companies perform credit analyses and express their conclusions by a system of ratings. The three commercial rating companies in the United States are (1) Moody's Investors Service (Moody's), (2) Standard & Poor's Corporation (S&P), and (3) FitchRatings (Fitch).

In all systems, the term **high grade** means low credit risk, or conversely, high probability of future payments. The highest-grade bonds are designated by Moody's by the symbol Aaa, and by S&P and Fitch by the symbol AAA. The next highest grade is denoted by the symbol Aa (Moody's) or AA (S&P and Fitch); for the third grade, all rating systems use A. The next three grades are Baa or BBB, Ba or BB, and B, respectively. There are also C grades. Moody's uses 1, 2, or 3 to provide a narrower credit quality breakdown within each class, and S&P and Fitch use plus and minus signs for the same purpose.

Bonds rated triple A (AAA or Aaa) are said to be *prime*; double A (AA or Aa) are of *high quality*; single A issues are called *upper medium grade*, and triple B are *medium grade*. Lower-rated bonds are said to have speculative elements or be distinctly speculative.

Bond issues that are assigned a rating in the top four categories are referred to as *investment-grade bonds*. Issues that carry a rating below the top four categories are referred to as *noninvestment-grade bonds*, or more popularly as *high-yield bonds* or *junk bonds*. Thus, the bond market can be divided into two sectors: the investment-grade and noninvestment-grade markets.

The spread between Treasury securities and non-Treasury securities that are identical in all respects except for credit quality is referred to as a **credit spread**. For example, Yahoo.com reported that the yield for corporate bonds with 10 years to maturity on May 20, 2008, was 5.14% for AAA rated bonds and 5.95% for AA rated bonds. The 10-year on-the-run Treasury rate was 3.78%. Hence, the yield spread for 10-year AAA corporate bonds

was 136 basis points (5.14% − 3.78%) and 217 basis points (5.95% − 3.78%) for AA rated corporate bonds. As can be seen, the lower the credit rating, the larger the yield spread.

Let us return to the interpretation if the yield spread on a non–U.S. Treasury security is used as the benchmark interest rate. Since the U.S. Treasury securities are viewed by market participants as free of credit risk, the yield spread to U.S. Treasury securities is reflecting credit risk, as well as the other risks to be discussed. In the search for a new benchmark interest rate, it is desirable to use a benchmark that has minimal credit risk. There are three possible candidates for the benchmark. The first is the yield on securities offered by government-sponsored enterprises. They are currently viewed as having a triple A rating (AAA or Aaa) and, therefore, have minimal credit risk. The second alternative is corporate bond issuers that have triple A ratings. As we will see in our discussion of liquidity, corporate bond issuers may not be a viable candidate to replace Treasury securities as the benchmark. The third candidate is one that comes from a market, the interest rate swap market. A measure that comes out of the swap market is called the swap rate, which many market participants use as a benchmark. We will discuss this later in this chapter.

Term to Maturity

The price of a financial asset will fluctuate over its life as yields in the market change. As we demonstrated, the volatility of a bond's price is dependent on its maturity. More specifically, with all other factors constant the longer the maturity of a bond, the greater the price volatility resulting from a change in market yields. The spread between any two maturity sectors of the market is called a **maturity spread** or **yield curve spread**.

Although this spread can be calculated for any sector of the market, it is most commonly calculated for the Treasury sector. For example, this spread on May 20, 2008, for the four on-the-run Treasury issues whose yields were given above are

two-year/five-year maturity spread	3.02%	− 2.34%	= 0.0068%	= 68 basis points
two-year/10-year maturity spread	3.78%	− 2.34%	= 0.0144%	= 144 basis points
two-year/30-year maturity spread	4.53%	− 2.34%	= 0.0219%	= 219 basis points
five-year/10-year maturity spread	3.78%	− 3.02%	= 0.0066%	= 66 basis points
five-year/30-year maturity spread	4.53%	− 3.02%	= 0.0151%	= 151 basis points
10-year/30-year maturity spread	4.53%	− 3.78%	= 0.0075%	= 75 basis points

Note that although the maturity spread is the difference between the yield on the longer maturity and the yield on the shorter maturity, the market convention is to refer to the maturity spread as the "shorter maturity/longer maturity."

There are other market conventions when referring to maturity spreads. For example, some market participants refer to the two-year/10-year maturity spread as the "short-end of the yield curve" and the 10-year/30-year maturity spread as the "long-end of the yield curve."

In the corporate bond market, a maturity spread can be similarly calculated but the issues must have the same credit quality. For example, on May 20, 2009, Yahoo.com reported the following for A rated corporate bonds:[5]

---

[5] Yahoo.com did not report any yield for 30-year A rated corporate bonds. This is because there are not many issues with this maturity outstanding.

two-year A corporate bonds 3.82%
five-year A corporate bonds 4.87%
10-year A corporate bonds 5.39%

The maturity spreads are then

two-year/five-year maturity spread	4.87% − 3.82%	= 0.0105%	=	105 basis points
two-year/10-year maturity spread	5.39% − 3.82%	= 0.0157%	=	157 basis points
five-year/10-year maturity spread	5.39% − 4.87%	= 0.0052%	=	52 basis points

The relationship between the yields on comparable securities but different maturities is called the **term structure of interest rates**. The term-to-maturity topic is of such importance that we devote the entire next chapter to it.

Inclusion of Options

It is not uncommon for a bond issue to include a provision that gives either the bondholder and/or the issuer an option to take some action against the other party. An option that is included in a bond issue is referred to as an **embedded option**.

The most common type of option in a bond issue is a **call provision**. This provision grants the issuer the right to retire the debt, fully or partially, before the scheduled maturity date. The inclusion of a call feature benefits issuers by allowing them to replace an old bond issue with a lower interest cost issue should interest rates in the market decline. Effectively, a call provision allows the issuer to alter the maturity of a bond. A call provision is detrimental to the bondholder because the bondholder will be uncertain about maturity and might have to reinvest the proceeds received at a lower interest rate if the bond is called and the bondholder wants to keep his or her funds in issues of similar risk of default.

An issue may also include a provision that allows the bondholder to change the maturity of a bond. An issue with a **put provision** grants the bondholder the right to sell the issue back to the issuer at par value on designated dates. Here, the advantage to the investor is that, if interest rates rise after the issue date and result in a price that is less than the par value, the investor can force the issuer to redeem the bond at par value.

A **convertible bond** is an issue giving the bondholder the right to exchange the bond for a specified number of shares of common stock. This feature allows the bondholder to take advantage of favorable movements in the price of the issuer's common stock.

The presence of these embedded options has an effect on the spread of an issue relative to a Treasury security and the spread relative to otherwise comparable issues that do not have an embedded option. In general, market participants will require a larger spread over a comparable Treasury security for an issue with an embedded option that is favorable to the issuer (e.g., a call option) than for an issue without such an option. In contrast, market participants will require a smaller spread over a comparable Treasury security for an issue with an embedded option that is favorable to the investor (for example, put option and conversion option). In fact, for a bond with an option that is favorable to an investor, the interest rate on an issue may be less than that on a comparable Treasury security!

A major part of the bond market is the mortgage market. There is a wide range of mortgage-backed securities. But these securities expose an investor to a form of call risk called "prepayment risk." Consequently, a yield spread between a mortgage-backed security and a comparable on-the-run Treasury security reflects this call risk. To see this, consider a basic mortgage-backed security called a Ginnie Mae passthrough

security. This security is backed by the full faith and credit of the U.S. government. Consequently, the yield spread between a Ginnie Mae passthrough security and a comparable Treasury security is not due to credit risk. Rather, it is primarily due to call risk.

Taxability of Interest

Unless exempted under the federal income tax code, interest income is taxable at the federal level. In addition to federal income taxes, there may be state and local taxes on interest income.

The federal tax code specifically exempts the interest income from qualified municipal bond issues from taxation at the federal level. **Municipal bonds** are securities issued by state and local governments and by their creations, such as "authorities" and special districts. The large majority of outstanding municipal bonds are tax-exempt securities. Because of the tax-exempt feature of municipal bonds, the yield on municipal bonds is less than that on Treasuries with the same maturity. The difference in yield between tax-exempt securities and Treasury securities is typically measured not in basis points but in percentage terms. More specifically, it is measured as the percentage of the yield on a tax-exempt security relative to a comparable Treasury security.

The yield on a taxable bond issue after federal income taxes are paid is equal to

$$\text{After-tax yield} = \text{Pretax yield} \times (1 - \text{Marginal tax rate})$$

For example, suppose a taxable bond issue offers a yield of 4% and is acquired by an investor facing a marginal tax rate of 35%. The after-tax yield would then be

$$\text{After-tax yield} = 0.04 \times (1 - 0.35) = 0.026 = 2.6\%$$

Alternatively, we can determine the yield that must be offered on a taxable bond issue to give the same after-tax yield as a tax-exempt issue. This yield is called the equivalent taxable yield and is determined as follows:

$$\text{Equivalent taxable yield} = \frac{\text{Tax-exempt yield}}{(1 - \text{Marginal tax rate})}$$

For example, consider an investor facing a 35% marginal tax rate who purchases a tax-exempt issue with a yield of 2.6%. The equivalent taxable yield is then

$$\text{Equivalent taxable yield} = \frac{0.026}{(1 - 0.35)} = 0.04 = 4\%$$

Notice that the lower the marginal tax rate, the lower the equivalent taxable yield. Thus, in our previous example, if the marginal tax rate is 25% rather than 35%, the equivalent taxable yield would be 3.47% rather than 4%, as shown below:

$$\text{Equivalent taxable yield} = \frac{0.026}{(1 - 0.25)} = 0.0347 = 3.47\%$$

State and local governments may tax interest income on bond issues that are exempt from federal income taxes. Some municipalities exempt interest income paid on all municipal issues from taxation; others do not. Some states exempt interest income from bonds issued by municipalities within the state but tax the interest income from bonds issued by municipalities outside of the state. The implication is that two municipal securities of the same quality rating and the same maturity may trade at some spread because of different tax policies and

hence, the relative demand for bonds of municipalities in different states. For example, in a high income tax state such as New York, the demand for bonds of municipalities will drive down their yield relative to municipalities in a low or zero income tax state such as Florida.

Municipalities are not permitted to tax the interest income from securities issued by the U.S. Treasury. Thus, part of the spread between Treasury securities and taxable non-Treasury securities of the same maturity reflects the value of the exemption from state and local taxes.

Expected Liquidity of an Issue

Bonds trade with different degrees of liquidity. The greater the expected liquidity with which an issue will trade, the lower the yield that investors would require. As noted earlier, Treasury securities are the most liquid securities in the world. The lower yield offered on Treasury securities relative to non-Treasury securities reflects, to a significant extent, the difference in liquidity. Even within the Treasury market, some differences in liquidity occur, because on-the-run issues have greater liquidity than off-the-run issues.

An important factor that affects the liquidity of an issue is the size of the issue. One of the reasons that U.S. Treasury securities are highly liquid is the large size of each individual issue. Government-sponsored enterprises in recent years have dramatically increased their issuance size. They have done this in their attempt to have the yield on their securities become the benchmark interest rates in the U.S. economy. As we noted earlier, the triple A credit rating of government-sponsored enterprises makes the yield on their securities an attractive candidate to replace the yield on Treasury securities as the benchmark interest rates. The large size and therefore liquidity of the securities issued by government-sponsored enterprises increase their attractiveness as an alternative benchmark. We also noted earlier that the yield on corporate bond issues rated triple A may make them a potential alternative benchmark. However, the size of each issue for the various triple A corporate bond issuers is considerably smaller than that of the securities issued by government-sponsored enterprises. For this reason, triple A corporate bond issuers may not be received by market participants as a benchmark alternative.

Swap Rate Yield Curve

The information that is obtained from the interest rate swap market is also used as an interest rate benchmark. We will briefly explain how and why here.

In a generic interest rate swap, the parties exchange interest rate payments on specified dates: One party pays a fixed rate and the other party a floating rate over the life of the swap. In a typical swap, the floating rate is based on a reference rate and the reference rate is typically the London Interbank Offered Rate (LIBOR). LIBOR is the interest rate at which prime banks in London pay other prime banks on U.S. dollar certificates of deposit.

The fixed interest rate that is paid by the fixed-rate counterparty is called the swap rate. Dealers in the swap market quote swap rates for different maturities. The relationship between the swap rate and maturity of a swap is called the **swap rate yield curve**, more commonly referred to as the **swap curve**. Because the reference rate is typically LIBOR, the swap curve is also called the **LIBOR curve**.

There is a swap curve for most countries. Table 3 shows the U.S. swap curve, Euro swap curve, and U.K. swap curve on October 16, 2007. For Euro interest rate swaps, the reference rate is the Euro Interbank Offered Rate (Euribor), which is the rate at which bank

Years to Maturity	Swap rate (%) for U.S.	Euro	U.K.
1	4.96	4.72	6.19
2	4.85	4.65	5.97
3	4.90	4.64	5.91
5	5.12	4.65	5.80
7	5.18	4.69	5.70
10	5.33	4.77	5.56
12	5.39	4.82	5.50
15	5.46	4.88	5.40
20	5.52	4.92	5.27
15	5.54	4.92	5.14
30	5.54	4.90	5.04

Note: *The rates in this table were obtained from CLP Structured Finance's website, http://www.swap-rates.com.*

deposits in countries that have adopted the euro currency and are member states of the European Union[6] are offered by one prime bank to another prime bank.

The swap curve is used as a benchmark in many countries outside the United States. Unlike a country's government bond yield curve, however, the swap curve is not a default-free yield curve. Instead, it reflects the credit risk of the counterparty to an interest rate swap. Since the counterparty to an interest rate swap is typically a bank-related entity, the swap curve reflects the average credit risk of representative banks that provide interest rate swaps. More specifically, a swap curve is viewed as the *interbank yield curve.* It is also referred to as the *AA rated yield curve* because the banks that borrow money from each other at LIBOR have credit ratings of Aa/AA or above. In addition, the swap curve reflects liquidity risk. However, in recent years the liquidity of the interest rate swap has increased to the point where it is now a more liquid market than the market for some government bonds.

One would expect that if a country has a government bond market, the yields in that market would be the best benchmark. That is not necessarily the case. There are several advantages of using a swap curve over a country's government securities yield curve.[7]

First, there may be technical reasons why within a government bond market some of the interest rates may not be representative of the true interest rate but instead be biased by some technical or regulatory factor unique to that market. For example, market participants may need to cover a short position in a particular government bond and the actions to cover a short position would push up the demand for that particular government bond and drive down its yield. In the swap market, there is nothing that has to be delivered so technical market factors have less of an impact. Also, there may be government bonds selling above or below their par value. Moreover, as explained earlier, government tax authorities might tax such bonds differently if they are purchased and held to maturity. As a

---

[6] The euro is the official currency of many European Union member states. Countries where the euro is adopted as the official currency are commonly referred to as part of the Eurozone, Euroland, or Euro area.

[7] For a further discussion, see Uri Ron, "A Practical Guide to Swap Curve Construction," Chapter 6 in Frank J. Fabozzi (ed.), *Interest Rate, Term Structure, and Valuation Modeling* (New York: John Wiley & Sons, 2002).

result, the yields at which these bonds trade in the marketplace will reflect any tax advantage or disadvantage. Although it may be difficult to appreciate these factors at this point in your study of financial markets, the key is that the observed interest rate on government securities may not reflect the true interest rate due to these factors. This is not the case for swap rates. There is no regulation of this market and hence swap rates represent true interest rates. However, remember that swap rates do reflect credit risk and liquidity risk.

Second, to create a representative government bond yield curve, a large number of maturities must be available. However, in most government bond markets, securities with only a few maturities are issued. For example, the U.S. government issues only four securities with a maturity of two years or more (two, five, 10, and 30 years). While there are a good number of off-the-run issues available from which to construct a government bond yield curve, the yields on such issues may not be true interest rates for the reasons noted previously. In contrast, in the swap market, a wide range of maturities are quoted, as can be seen in Table 1. In fact, in the United States, there was a suspension in the issuance of 30-year bonds, and as a result, 30-year Treasury rates were unavailable. Yet, swap dealers quoted 30-year swap rates.

Finally, the ability to compare government yields across countries is difficult because there are differences in the credit risk for every country. In contrast, as explained earlier, the swap curve is an interbank yield curve and thereby makes cross-country comparisons of benchmark interest rates easier.

Key Points That You Should Understand Before Proceeding

1. Historically, the on-the-run Treasury security of a given maturity is the base or benchmark interest rate.
2. With the decline in the issuance of Treasury securities by the U.S. Department of the Treasury, there is now a search for an alternative interest rate benchmark.
3. Non-Treasury securities will trade at a spread (measured in basis points) relative to an on-the-run Treasury security.
4. The factors that affect yield spreads in the bond market are the type of issuer, credit risk, term to maturity, the presence of embedded options, the tax treatment of interest income, and liquidity.
5. The credit risk of an issuer is gauged by ratings assigned by commercial rating companies (Moody's, S&P, and Fitch).
6. Embedded options can either reduce the spread to Treasuries, if the option benefits the bondholder, or increase the spread to Treasuries, if the option benefits the issuer.
7. The interest income from municipal securities is generally exempt from federal income taxation, and as a result municipals offer a lower yield than Treasury securities.

SUMMARY

Two important theories of the determination of the level of the interest rate are Fisher's theory, adapted to the loanable funds model, and Keynes's liquidity preference theory. Fisher considers the reasons for saving to be the marginal rate of time preference, which is the willingness to forgo current consumption for enhanced future consumption, the level of income, and the

rate of interest. The demand for borrowing arises because of firms' investment opportunities. The marginal productivity of capital is the gain from additional investment and it is negatively related to the interest rate. The equilibrium rate of interest is determined by the interaction of the demand for and supply of savings. In the presence of inflation, the equilibrium rate is composed of both the real rate of interest and a premium for expected inflation.

Keynes's theory emphasizes the role of liquid balances or money in transaction and states that the interest rate is determined in the money market. Demand for money reflects income and the level of prices for goods and services, and it is negatively related to the level of interest rates. The supply of money, in this theory, is controlled by the central bank, which can, by changing the money supply, affect the rate of interest. In fact, changes in the money supply can have three possibly conflicting effects on the level of the interest rate, depending on the economy's level of output and employment. Those effects are the liquidity effect, the income effect, and the price expectations effect.

The basic features of a bond include the coupon rate, the term to maturity, and the principal or par value. The yield to maturity of a bond is the interest rate that makes the total present value of the bond's cash flow equal to its market price.

Every economy operates with not just one interest rate but rather with a structure of interest rates. The yield spread is the difference between the yield on any two bonds and reflects the difference in their risks. A country's government bond market can be used as the base interest rate for that country. The base interest rate is equal to the real rate of interest plus the expected rate of inflation. The risk premium is the yield spread between a non-government security and a comparable government security. The factors that affect the yield spread include: (1) the type of issuer (agency, corporation, municipality), (2) the issuer's perceived creditworthiness as measured by the rating system of commercial rating companies, (3) the term or maturity of the instrument, (4) the embedded options in a bond issue (e.g., call, put, or conversion provisions), (5) the taxability of interest income at the federal and municipal levels, and (6) the expected liquidity of the issue.

The swap rate yield curve also provides information about interest rates in a country. The swap rate yield curve or simply swap curve is not a default-free yield curve but rather reflects interbank credit risk. In many countries, market participants use the country's swap curve as the benchmark interest rates rather than the country's government bond yield curve.

KEY TERMS

- Base interest rate (benchmark interest rate)
- Bond-equivalent yield basis
- Call provision
- Convertible bond
- Credit risk
- Credit spread
- Default risk
- Embedded option
- Equivalent taxable yield
- Fisher's Law
- Fisher's theory of interest
- High-grade
- Income effect
- Intermarket sector spread
- Intramarket sector spread
- LIBOR curve
- Liquidity effect
- Liquidity preference theory of interest
- Loanable funds theory of interest rates
- Marginal productivity of capital
- Marginal rate of time preference
- Market sectors
- Maturity

- Maturity spread
- Municipal bonds
- Nominal rate of interest
- Off-the-run Treasury issue
- On-the-run Treasury issue (current coupon issue)
- Price expectations effect
- Principal value (principal, par value, maturity value, redemption value, face value)
- Put provision
- Real rate of interest
- Risk premium
- Swap curve
- Swap rate
- Swap rate yield curve
- Term to maturity
- Term structure of interest rates
- Treasury security (Treasuries)
- Yield curve spread
- Yield spread (spread)
- Yield to maturity

QUESTIONS

1. Explain what these terms mean in Fisher's theory of interest rates:
   a. *marginal rate of time preference*
   b. *marginal productivity of capital*
   c. *equilibrium interest rate*
2. a. How does the loanable funds theory expand Fisher's theory of interest rate determination?
   b. How does a change in the government's deficit affect the equilibrium rate in the loanable funds theory?
3. a. How do the assets, money, and bonds differ in Keynes's liquidity preference theory?
   b. How does a change in income affect the equilibrium level of the interest rate in Keynes's theory?
   c. How does a change in the money supply affect that rate?
4. a. Explain these terms: the *liquidity effect*; the *income effect*; and the *price expectations effect.*
   b. How is it determined which of these three effects on interest rates an increase in the money supply will have?
5. Consider three bonds, all with a par value of $1,000:

Bond	Coupon Rate	Market Price
A	8%	$1,100
B	7	900
C	9	1,000

   a. What is the yield to maturity of bond C?
   b. Is the yield to maturity of bond A greater than or less than 8%?
   c. Is the yield to maturity of bond B greater than or less than 9%?
6. Consider the following bond: 19 years to maturity, 11% coupon rate, pays interest semiannually, $1,000 par value. Suppose that the market price of this bond is $1,233.64. Given the information below, determine the yield to maturity for this bond on a bond-equivalent yield basis.

Interest Rate Tried	Total Present Value of Cash Flow
3.00%	$1,562.32
3.25	1,486.96
3.50	1,416.82
3.75	1,351.46
4.00	1,290.52
4.25	1,233.64

7. a. Show the cash flows for the two bonds below, each of which has a par value of $1,000 and pays interest semiannually:

Bond	Coupon Rate	Years to Maturity	Price
W	7%	5	$884.10
Y	9	4	967.21

   b. Calculate the yield to maturity for the two bonds.
8. a. What is the credit risk associated with a U.S. Treasury security?
   b. Why is the Treasury yield considered the base interest rate?
   c. What is meant by *on-the-run Treasuries*?
   d. What is meant by *off-the-run Treasuries*?

9. What does the yield spread between the off-the-run Treasury issue and the on-the-run Treasury issue reflect?
10. On October 19, 2007, the yield for three on-the-run issues as reported by Yahoo.com was

    two-year Treasury 3.90%
    five-year Treasury 4.14%
    10-year Treasury 4.49%

    Compute all of the maturity spreads.
11. Yahoo.com reported that the yield for corporate bonds with 10 years to maturity on October 19, 2007, was 5.39% for AAA rated bonds and 5.66% for AA rated bonds. At the time, the 10-year on-the-run Treasury rate was 4.49%. Compute the two credit spreads.
12. In the May 29, 1992, *Weekly Market Update,* published by Goldman Sachs & Co., the following information was reported in various exhibits for certain corporate bonds as of the close of business Thursday, May 28, 1992:

Issuer	Rating	Yield	Spread	Treasury Benchmark
General Electric Capital Co.	Triple A	7.87%	50	10
Mobil Corp.	Double A	7.77	40	10
Southern Bell Tel & Teleg	Triple A	8.60	72	30
Bell Tel Co Pa	Double A	8.66	78	30
AMR Corp	Triple B	9.43	155	30

    a. What is meant by *rating*?
    b. Which of the five bonds has the greatest credit risk?
    c. What is meant by *spread*?
    d. What is meant by *Treasury benchmark*?
    e. Explain how each of the spreads reported above was determined.
    f. Why does each spread reported above reflect a risk premium?
13. For the corporate bond issues reported in the previous question, answer the following questions:

    a. Should a triple A rated bond issue offer a higher or lower yield than a double A rated bond issue of the same maturity?
    b. What is the spread between the General Electric Capital Company issue and the Mobil Corporation issue?
    c. Is the spread reported in (b) consistent with your answer to (a)?
    d. The yield spread between these two bond issues reflects more than just credit risk. What other factors would the spread reflect?
    e. The Mobil Corporation issue is not callable. However, the General Electric Capital Company issue is callable. How does this information help you in understanding the spread between these two issues?
14. For the corporate bond issues reported in question 12, answer the following questions:

    a. What is the yield spread between the Southern Bell Telephone and Telegraph bond issue and the Bell Telephone Company (Pennsylvania) bond issue?
    b. The Southern Bell Telephone and Telegraph bond issue is not callable but the Bell Telephone Company (Pennsylvania) bond issue is callable. What does the yield spread in (a) reflect?
    c. AMR Corporation is the parent company of American Airlines and is therefore classified in the transportation industry. The bond issue cited is not callable. What is the yield spread between the AMR Corporation and Southern Bell Telephone and Telegraph bond issue, and what does this spread reflect?
15. In the May 29, 1992, *Weekly Market Update,* published by Goldman Sachs & Co., the following information was reported in an exhibit for high-grade, tax-exempt securities as of the close of business Thursday, May 28, 1992:

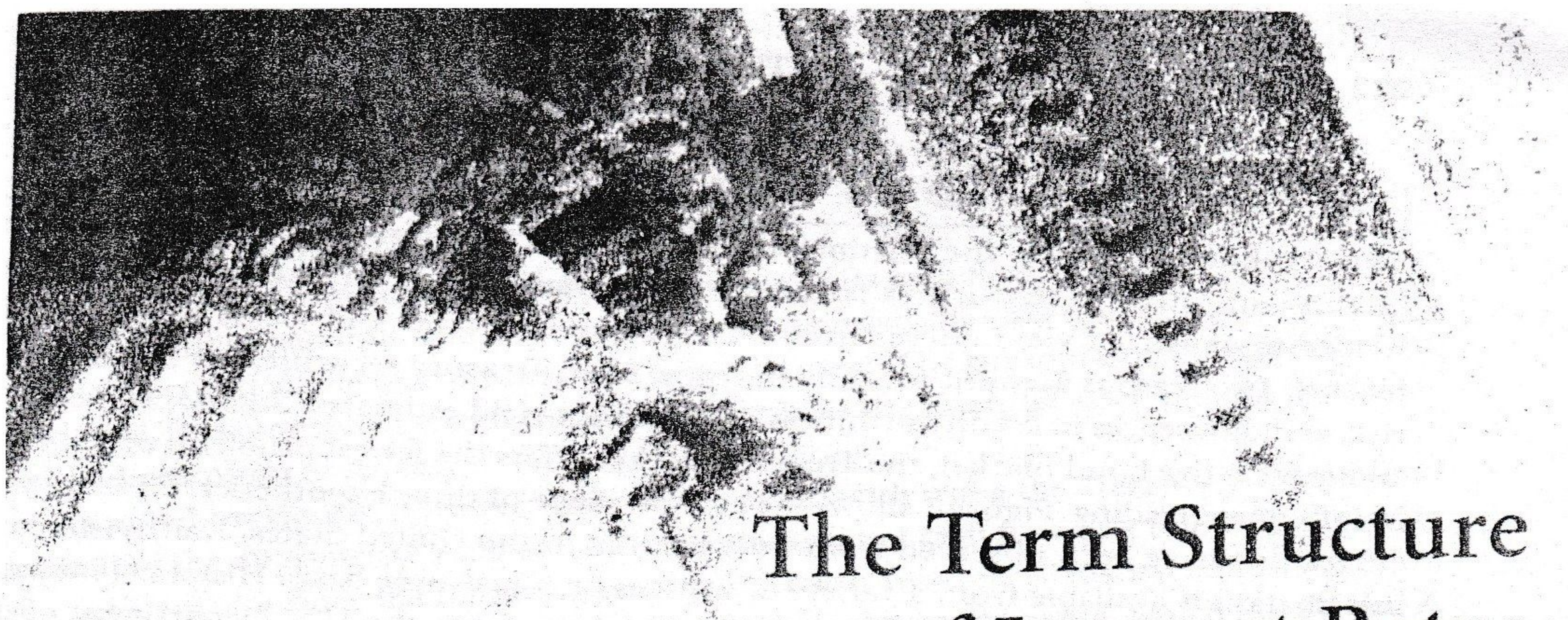
Maturity	Yield	Yield as a Percent of Treasury Yield
1-year	3.20%	76.5%
3-year	4.65	80.4
5-year	5.10	76.4
10-year	5.80	78.7
30-year	6.50	82.5

a. What is meant by a *tax-exempt security*?
b. What is meant by a *high-grade* issue?
c. Why is the yield on a tax-exempt security less than the yield on a Treasury security of the same maturity?
d. What is meant by the *equivalent taxable yield*?
e. Also reported in the same issue of the Goldman Sachs report is information on "Intra-market Yield Spreads." What is an intra-market yield spread?

16. a. What is meant by an *embedded option* in a bond?
b. Give three examples of an embedded option that might be included in a bond issue.
c. Does an embedded option increase or decrease the risk premium relative to the base interest rate?

17. a. What is meant by the swap rate?
b. What is meant by the swap curve?
c. Explain whether you agree or disagree with the following statement: "A country's swap curve is default-free yield curve."

18. Why do market participants in some countries prefer to use the swap curve rather than the government bond yield curve?



The Term Structure of Interest Rates

LEARNING OBJECTIVES

After reading this chapter, you will understand

- what is meant by the term structure of interest rates
- what the yield curve is
- the different shapes that the term structure can take
- what is meant by a spot rate and a spot rate curve
- how a theoretical spot rate curve can be determined from the Treasury yield curve
- what is meant by an implicit forward rate and how it can be calculated
- how long-term rates are related to the current short-term rate and short-term forward rates
- the different theories about the determinants of the shape of the term structure: pure expectations theory, the liquidity theory, the preferred habitat theory, and the market segmentation theory
- the risks associated with investing in bonds when interest rates change—price risk and reinvestment
- the main economic influences on the shape of the Treasury yield curve

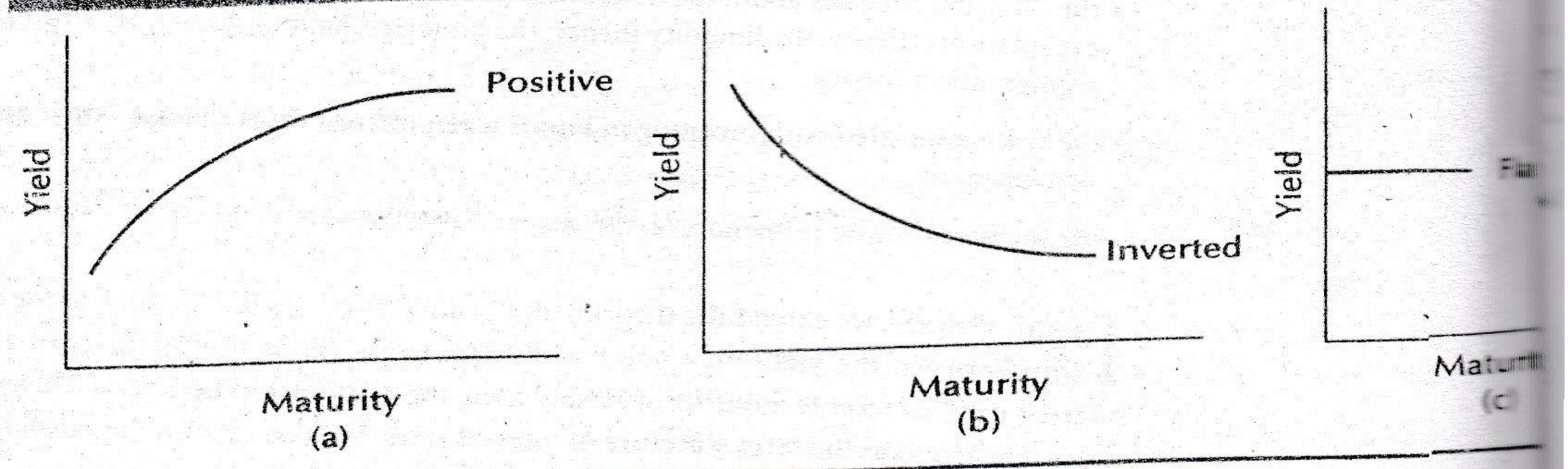
In this chapter, we extend the theories and principles of the last chapter to the relationship between the yield on a bond and its maturity. Since the maturity of a bond is referred to as its *term to maturity* or simply *term*, the relationship between yield and maturity is referred to as the **term structure of interest rates**. We also explain the various theories about the determinants of the term structure of interest rates.

From Chapter 11 of *Foundations of Financial Markets and Institutions*, 4/e. Frank J. Fabozzi. Franco Modigliani. Frank J. Jones.
Copyright © 2010 by Pearson Prentice Hall. All rights reserved.

THE YIELD CURVE AND THE TERM STRUCTURE

The graphic that depicts the relationship between the yield on bonds of the same credit quality but different maturities is known as the **yield curve**. Market participants have tended to construct yield curves from observations of prices and yields in the Treasury market. Two reasons account for this tendency. First, Treasury securities are free of default risk, and differences in creditworthiness do not affect yield estimates. Second, as the largest and most active bond market, the Treasury market offers the fewest problems of illiquidity or infrequent trading. Figure 1 shows the general shape of three hypothetical Treasury yield curves that have been observed from time to time in the United States. Daily yield curve information is available from a variety of sources on a real-time basis. Historical information about daily yield curves from 1990 is obtained from the U.S. Department of the Treasury's website.[1]

From a practical viewpoint, as we explained in the previous chapter, the Treasury yield curve functions mainly as a benchmark for pricing bonds and setting yields in many other sectors of the debt market—bank loans, mortgages, corporate debt, and international bonds. More recently market participants have come to realize that the traditionally constructed Treasury yield curve is an unsatisfactory measure of the relation between required yield and maturity. The key reason is that securities with the same maturity may actually provide different yields. As we will explain below, this phenomenon reflects the role and impact of differences in the bonds' coupon rates. Hence, it is necessary to develop more accurate and reliable estimates of the Treasury yield curve. In what follows, we will show the problems posed by traditional approaches to the Treasury yield curve and offer an innovative and increasingly popular approach to building a yield curve. The approach consists of identifying yields that apply to zero-coupon bonds and, therefore, eliminates the problem of nonuniqueness in the yield–maturity relationship.



[1] http://www.ustreas.gov/offices/domestic-finance/debt-management/interest-rate/yield_historical_main.shtml.

Using the Yield Curve to Price a Bond

The price of any financial asset is the present value of its cash flow. However, in our illustrations and our discussion to this point, we have assumed that one interest rate should be used to discount all the cash flows from a financial asset. The appropriate interest rate is the yield on a Treasury security with the same maturity as the financial asset, plus an appropriate yield premium or spread.

As noted previously, however, there is a problem with using the Treasury yield curve to determine the appropriate yield at which to discount the cash flow of a bond. To illustrate this problem, consider the following two hypothetical five-year Treasury bonds, A and B. The difference between these two Treasury bonds is the coupon rate, which is 12% for A and 3% for B. The cash flow for these two bonds per $100 of par value for the 10 six-month periods to maturity would be:

Period	Cash Flow for A	Cash Flow for B
1–9	$ 6.00	$ 1.50
10	106.00	101.50

Because of the different cash flow patterns, it is not appropriate to use the same interest rate to discount all cash flows. Instead, each cash flow should be discounted at a unique interest rate that is appropriate for the time period in which the cash flow will be received. But what should be the interest rate for each period?

The correct way to think about bonds A and B is not as bonds but as packages of cash flows. More specifically, they are packages of zero-coupon instruments. A **zero-coupon instrument** is one that is purchased at an amount below its maturity value and that pays no interest periodically. Instead, the interest is earned at the maturity date when the investor receives the maturity or principal value. Thus, the interest earned is the difference between the maturity value and the price paid. For example, bond A can be viewed as 10 zero-coupon instruments: one with a maturity value of $6 maturing six months from now; a second with a maturity value of $6 maturing one year from now; a third with a maturity value of $6 maturing 1.5 years from now, and so on. The final zero-coupon instrument matures 10 six-month periods from now and has a maturity value of $106.

Likewise, bond B can be viewed as 10 zero-coupon instruments: one with a maturity value of $1.50 maturing six months from now; one with a maturity value of $1.50 maturing one year from now; one with a maturity value of $1.50 maturing 1.5 years from now, and so on. The final zero-coupon instrument matures 10 six-month periods from now and has a maturity value of $101.50. Obviously, in the case of each coupon bond, the value or price of the bond is equal to the total value of its component zero-coupon instruments.

In general, any bond can be viewed as a package of zero-coupon instruments. That is, each zero-coupon instrument in the package has a maturity equal to its coupon payment date or, in the case of the principal, the maturity date. The value of the bond should equal the value of all the component zero-coupon instruments. If this does not hold, it is possible for a market participant to generate riskless profits. Because no one can pass up riskless and certain profits, the market must drive these two prices to equality, and our discussion here assumes that equality.

To determine the value of each zero-coupon instrument, it is necessary to know the yield on a zero-coupon Treasury with that same maturity. This yield is called the **spot rate**, and the

graphical depiction of the relationship between the spot rate and its maturity is called the spot rate curve. Below we will see how to derive this curve from theoretical considerations of the yields of the actually traded Treasury securities. Such a curve is called a *theoretical spot rate curve.*

Constructing the Theoretical Spot Rate Curve

In this section, we explain how the theoretical spot rate curve is constructed from the yield curve that is based on the observed yields of Treasury bills and Treasury coupon securities. The process of creating a theoretical spot rate curve in this way is called **bootstrapping.**[2] To explain this process, we will use the data for the hypothetical price, annualized yield (yield to maturity), and maturity of the 20 Treasury securities shown in Table 1. (In practice, all the coupon rates are estimated so that the price of each bond is par.)

Table 1 Maturity and Yield to Maturity for 20 Hypothetical Treasury Securities

Maturity (years)	Coupon Rate	Yield to Maturity	Price
0.50	0.0000	0.0800	$ 96.15
1.00	0.0000	0.0830	92.19
1.50	0.0850	0.0890	99.45
2.00	0.0900	0.0920	99.64
4.00	0.1000	0.1040	98.72
4.50	0.1150	0.1060	103.16
5.00	0.0875	0.1080	92.24
5.50	0.1050	0.1090	98.38
6.00	0.1100	0.1120	99.14
2.50	0.1100	0.0940	103.49
3.00	0.0950	0.0970	99.49
3.50	0.1000	0.1000	100.00
6.50	0.0850	0.1140	86.94
7.00	0.0825	0.1160	84.24
7.50	0.1100	0.1180	96.09
8.00	0.0650	0.1190	72.62
8.50	0.0875	0.1200	82.97
9.00	0.1300	0.1220	104.30
9.50	0.1150	0.1240	95.06
10.00	0.1250	0.1250	100.00

[2] In practice, the Treasury securities that are used to construct the theoretical spot rate curve are the most recently auctioned Treasury securities of a given maturity. Such issues are referred to as the *on-the-run Treasury issues.* There are actual zero-coupon Treasury securities with a maturity greater than one year that are outstanding in the market. These securities are not issued by the U.S. Treasury but are created by certain market participants from actual coupon Treasury securities. It would seem logical that the observed yield on zero-coupon Treasury securities can be used to construct an actual spot rate curve, but there are problems with this approach. First, the liquidity of these securities is not as great as that of the coupon Treasury market. Second, there are maturity sectors of the zero-coupon Treasury market that attract specific investors who may be willing to trade off yield in exchange for an attractive feature associated with that particular maturity sector, thereby distorting the term structure relationship.

Throughout the analysis and illustrations to come, it is important to remember that the basic principle underlying bootstrapping is that the value of the Treasury coupon security should be equal to the value of the package of zero-coupon Treasury securities that duplicates the coupon bond's cash flow.

Consider the six-month Treasury bill in Table 1. As we explained in the previous chapter, a Treasury bill is a zero-coupon instrument. Therefore, its annualized yield of 8% is equal to the spot rate. Similarly, for the one-year Treasury, the cited yield of 8.3% is the one-year spot rate.[3] Given these two spot rates, we can compute the spot rate for a theoretical 1.5-year zero-coupon Treasury. The price of a theoretical 1.5-year zero-coupon Treasury should equal the present value of the three cash flows from an actual 1.5-year coupon Treasury, where the yield used for discounting is the spot rate corresponding to the cash flow. Using $100 as par, the cash flow for the 1.5-year Treasury with the 8.5% coupon rate is

$$\begin{aligned} &0.5 \text{ years} && 0.085 \times \$100 \times 0.5 && = \$4.25 \\ &1.0 \text{ years} && 0.085 \times \$100 \times 0.5 && = \$4.25 \\ &1.5 \text{ years} && 0.085 \times \$100 \times 0.5 + 100 && = \$104.25 \end{aligned}$$

The present value of the cash flow is then

$$\frac{4.25}{(1+z_1)^1} + \frac{4.25}{(1+z_2)^2} + \frac{104.25}{(1+z_3)^3}$$

where

z_1 = one-half the annualized six-month theoretical spot rate
z_2 = one-half the one-year theoretical spot rate
z_3 = one-half the 1.5-year theoretical spot rate

Since the six-month spot rate and one-year spot rate are 8.0% and 8.3%, respectively, we know these facts:

$$z_1 = 0.04 \text{ and } z_2 = 0.0415$$

We can compute the present value of the 1.5-year coupon Treasury security as

$$\frac{4.25}{(1.0400)^1} + \frac{4.25}{(1.0415)^2} + \frac{104.25}{(1+z_3)^3}$$

Since the price of the 1.5-year coupon Treasury security (from Table 1) is $99.45, the following relationship between market price and the present value of the cash flow must hold:

$$99.45 = \frac{4.25}{(1.0400)^1} + \frac{4.25}{(1.0415)^2} + \frac{104.25}{(1+z_3)^3}$$

We can solve for the theoretical 1.5-year spot rate as follows:

$$99.45 = 4.08654 + 3.91805 + \frac{104.25}{(1+z_3)^3}$$

$$91.44541 = \frac{104.25}{(1+z_3)^3}$$

[3] For zero-coupon securities such as Treasury bills, the price is less than par value.

$$(1 + z_3)^3 = 1.140024$$
$$z_3 = 0.04465$$

Doubling this yield, we obtain the bond equivalent yield of 0.0893 or 8.93%, which is the theoretical 1.5-year spot rate. That rate is the rate that the market would apply to a 1.5-year zero-coupon Treasury security if, in fact, such a security existed.

Given the theoretical 1.5-year spot rate, we can obtain the theoretical two-year spot rate. The cash flow for the two-year coupon Treasury in Table 1 is

0.5 years	$0.090 \times \$100 \times 0.5$	$= \$4.50$
1.0 years	$0.090 \times \$100 \times 0.5$	$= \$4.50$
1.5 years	$0.090 \times \$100 \times 0.5$	$= \$4.50$
2.0 years	$0.090 \times \$100 \times 0.5 + 100$	$= \$104.50$

The present value of the cash flow is then

$$\frac{4.50}{(1 + z_1)^1} + \frac{4.50}{(1 + z_2)^2} + \frac{4.50}{(1 + z_3)^3} + \frac{104.50}{(1 + z_4)^4}$$

where z_4 is one-half the two-year theoretical spot rate. Since the six-month spot rate, one-year spot rate, and 1.5-year spot rate are 8.0%, 8.3%, and 8.93%, respectively, then

$$z_1 = 0.04, z_2 = 0.0415, \text{ and } z_3 = 0.04465$$

Therefore, the present value of the two-year coupon Treasury security is

$$\frac{4.50}{(1.0400)^1} + \frac{4.50}{(1.0415)^2} + \frac{4.50}{(1.04465)^3} + \frac{104.50}{(1 + z_4)^4}$$

Since the price of the two-year coupon Treasury security is \$99.64, the following relationship must hold:

$$99.64 = \frac{4.50}{(1.0400)^1} + \frac{4.50}{(1.0415)^2} + \frac{4.50}{(1.04465)^3} + \frac{104.50}{(1 + z_4)^4}$$

We can solve for the theoretical two-year spot rate as follows:

$$99.64 = 4.32692 + 4.14853 + 3.94730 + \frac{104.25}{(1 + z_4)^4}$$

$$87.21725 = \frac{104.25}{(1 + z_4)^4}$$

$$(1 + z_4)^4 = 1.198158$$

$$z_4 = 0.046235$$

Doubling this yield, we obtain the theoretical two-year spot rate bond-equivalent yield of 9.247%.

One can follow this approach sequentially to derive the theoretical 2.5-year spot rate from the calculated values of z_1, z_2, z_3, and z_4 (the six-month, one-year, 1.5-year, and two-year rates), and the price and coupon of the bond with a maturity of 2.5 years. Furthermore, one could derive theoretical spot rates for the remaining 15 half-yearly rates. The spot rates thus obtained are shown in Table 2. They represent the term structure of interest rates for maturities up to 10 years, at the particular time to which the bond price quotations refer.

Maturity (years)	Yield to Maturity	Theoretical Spot Rate
0.50	0.0800	0.08000
1.00	0.0830	0.08300
1.50	0.0890	0.08930
2.00	0.0920	0.09247
2.50	0.0940	0.09468
3.00	0.0970	0.09787
3.50	0.1000	0.10129
4.00	0.1040	0.10592
4.50	0.1060	0.10850
5.00	0.1080	0.11021
5.50	0.1090	0.11175
6.00	0.1120	0.11584
6.50	0.1140	0.11744
7.00	0.1160	0.11991
7.50	0.1180	0.12405
8.00	0.1190	0.12278
8.50	0.1200	0.12546
9.00	0.1220	0.13152
9.50	0.1240	0.13577
10.00	0.1250	0.13623

Column 2 of Table 2 reproduces the calculated yield to maturity for the coupon issue listed in Table 1. A comparison of this column with the last column giving the yield to maturity of a zero-coupon bond is instructive, for it confirms that bonds of the same maturity may have different yields to maturity. That is, the yield of bonds of the same credit quality does not depend on their maturity alone. While the two columns do not change much at the beginning, they diverge more after the third year, and by the ninth year the zero-coupon yield is nearly 100 basis points higher than that of the same maturity with a coupon of 13% and selling at a premium.

Using Spot Rates to Value a Bond

Given the spot rates, the theoretical value of a bond can be calculated. This is done by discounting a cash flow for a given period by the corresponding spot rate for that period. This is illustrated in Table 3.

The bond in our illustration is a 10-year, 10% coupon Treasury bond. The second column of Table 3 shows the cash flow per $100 of par value for a 10% coupon bond. The third column shows the theoretical spot rates. The fourth column is simply one half of the annual spot rate of the previous column. The last column shows the present value of the cash flow in the second column when discounted at the semiannual spot rate. The value of this bond is the total present value, $85.35477.

Maturity (years)	Case Flow	Spot Rate	Semiannual Spot Rate	Present Value
0.5	5	0.08000	0.04000	4.8077
1.0	5	0.08300	0.04150	4.6095
1.5	5	0.08930	0.04465	4.3859
2.0	5	0.09247	0.04624	4.1730
2.5	5	0.09468	0.04734	3.9676
3.0	5	0.09787	0.04894	3.7539
3.5	5	0.10129	0.05065	3.5382
4.0	5	0.10592	0.05296	3.3088
4.5	5	0.10850	0.05425	3.1080
5.0	5	0.11021	0.05511	2.9242
5.5	5	0.11175	0.05588	2.7494
6.0	5	0.11584	0.05792	2.5441
6.5	5	0.11744	0.05872	2.3813
7.0	5	0.11991	0.05996	2.2128
7.5	5	0.12405	0.06203	2.0274
8.0	5	0.12278	0.06139	1.9274
8.5	5	0.12546	0.06273	1.7774
9.0	5	0.13152	0.06576	1.5889
9.5	5	0.13377	0.06689	1.4613
10.0	105	0.13623	0.06812	28.1079
Total				85.35477

Key Points That You Should Understand Before Proceeding

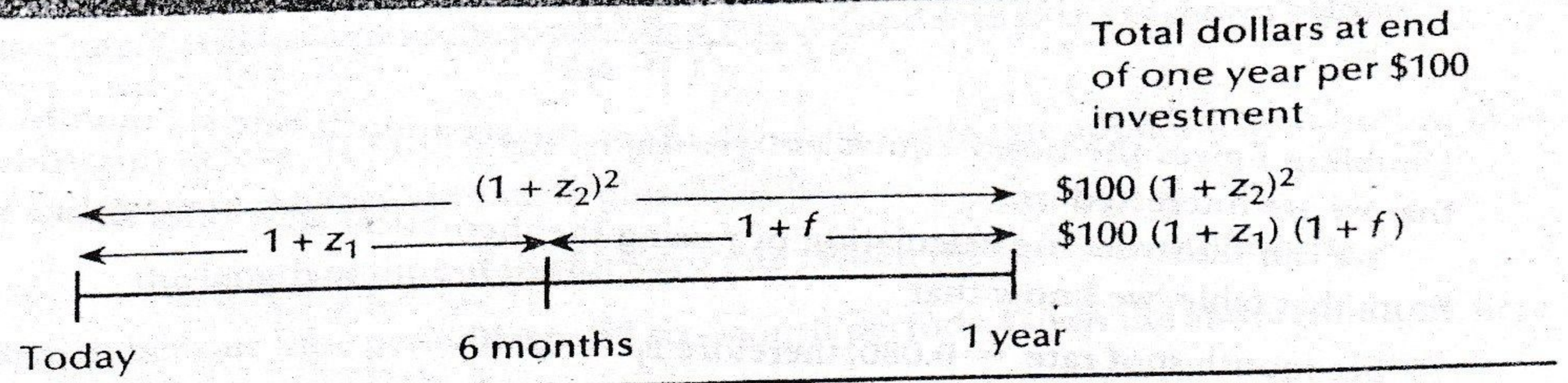
1. What the yield curve represents.
2. Why any financial asset can be viewed as a package of zero-coupon instruments.
3. What is meant by spot rates and how the theoretical spot rates can be constructed from a Treasury yield curve.
4. How to value a bond using spot rates.

FORWARD RATES

Thus far, we have seen that from the Treasury yield curve we can extrapolate the theoretical spot rates. In addition, we can extrapolate what some market participants refer to as the *market's consensus of future interest rates.* To see the importance of knowing the market's consensus for future interest rates, consider the following two investment alternatives for an investor who has a one-year investment horizon:

Alternative 1: Investor buys a one-year instrument.
Alternative 2: Investor buys a six-month instrument and when it matures in six months, the investor buys another six-month instrument.

Figure 2 Two Alternative One-Year Investments



With Alternative 1, the investor will realize the one-year spot rate and that rate is known with certainty. In contrast, with Alternative 2, the investor will realize the six-month spot rate, but the six-month rate six months from now is unknown. Therefore, for Alternative 2, the rate that will be earned over one year is not known with certainty. This is illustrated in Figure 2.

Suppose that this investor expected that six months from now the six-month rate will be higher than it is today. The investor might then feel Alternative 2 would be the better investment. However, this is not necessarily true. To understand why and to appreciate the need to understand why it is necessary to know what the market's consensus of future interest rates is, let's continue with our illustration.

The investor will be indifferent to the two alternatives if they produce the same total dollars over the one-year investment horizon. Given the one-year spot rate, there is some rate on a six-month instrument six months from now that will make the investor indifferent between the two alternatives. We will denote that rate by f.

The value of f can be readily determined given the one-year spot rate and the six-month spot rate. If an investor placed $100 in a one-year instrument (Alternative 1), the total dollars that will be generated at the end of one year is

$$\text{Total dollars at the end of year for Alternative 1} = \$100(1 + z_2)^2$$

where z_2 is the one-year spot rate. (Remember we are working in six-month periods, so the subscript 2 represents two six-month periods, or one year.)

The proceeds from investing at the six-month spot rate will generate the following total dollars at the end of six months:

$$\text{Total dollars at the end of six months for Alternative 2} = \$100(1 + z_1)$$

where z_1 is the six-month spot rate. If this amount is reinvested at the six-month rate six months from now, which we denoted f, then the total dollars at the end of one year would be

$$\text{Total dollars at the end of year for Alternative 2} = \$100(1 + z_1)(1 + f)$$

The investor will be indifferent between the two alternatives if the total dollars are the same. Setting the two equations for the total dollars at end of one year for the two alternatives equal we get

$$\$100 = (1 + z_2)^2 = \$100(1 + z_1)(1 + f)$$

Solving the preceding equation for f, we get

$$f = \frac{(1 + z_2)^2}{(1 + z_1)} - 1$$

Doubling f gives the bond-equivalent yield for the six-month rate six months from now that we are interested in.

We can illustrate the calculation of f using the theoretical spot rates shown in Table 2. From that table, we know that

Six-month spot rate = 0.080, therefore $z_1 = 0.0400$

One-year spot rate = 0.083, therefore $z_2 = 0.0415$

Substituting into the formula, we have

$$f = \frac{(1.0415)^2}{(1.0400)} - 1$$
$$= 0.043$$

Therefore, the forward rate on a six-month security, quoted on a bond-equivalent basis, is 8.6% (0.043 × 2).

Here is how we use this rate of 8.6%. If the six-month rate six months from now is less than 8.6%, then the total dollars at the end of one year would be higher by investing in the one-year instrument (Alternative 1). If the six-month rate six months from now is greater than 8.6%, then the total dollars at the end of one year would be higher by investing in the six-month instrument and reinvesting the proceeds six months from now at the six-month rate at the time (Alternative 2). Of course, if the six-month rate six months from now is 8.6%, the two alternatives give the same total dollars at the end of one year.

Now that we have the rate f in which we are interested and we know how that rate can be used, let's return to the question we posed at the outset. From Table 2, the six-month spot rate is 8%. Suppose that the investor expects that six months from now, the six month rate will be 8.2%. That is, the investor expects that the six-month rate will be higher than its current level. Should the investor select Alternative 2 because the six-month rate six months from now is expected to be higher? The answer is no. As we explained in the previous paragraph, if the rate is less than 8.6%, then Alternative 1 is the better alternative. Since this investor expects a rate of 8.2%, then he or she should select Alternative 1 despite the fact that he or she expects the six-month rate to be higher than it is today.

This is a somewhat surprising result for some investors. But the reason for this is that the market prices its expectations of future interest rates into the rates offered on investments with different maturities. This is why knowing the market's consensus of future interest rates is critical. The rate that we determined for f is the market's consensus for the six-month rate six months from now. A future interest rate calculated from either the spot rates or the yield curve is called a **forward rate** or an **implied forward rate.**

Similarly, borrowers need to understand what a forward rate is. For example, suppose a borrower must choose between a one-year loan and a series of two six-month loans. If the forward rate is less than the borrower's expectations of six-month rates six months from now, then the borrower will be better off with a one-year loan. If, instead, the borrower's expectations are that six-month rates six months from now will be less than the forward rate, the borrower will be better off by choosing a series of two six-month loans.

There are two elements to the forward rate. The first is when in the future the rate begins. The second is the length of time for the rate. For example, the two-year forward rate three years from now means a rate three years from now for a length of two years. The notation used for a forward rate, f, will have two subscripts—one before f and one after f as shown below:

$${}_tf_m$$

The subscript before f is t and is the length of time that the rate applies. The subscript after f is m and is when the forward rate begins. That is,

the length of time of the forward rate f when the forward rate begins

Remember our time periods are still six-month periods. Given the above notation, here is what the following mean:

Notation	Interpretation for the Forward Rate
${}_1f_{12}$	six-month (one-period) forward rate beginning six years (12 periods)from now
${}_2f_8$	one-year (two-period) forward rate beginning four years (8 periods)from now
${}_6f_4$	three-year (six-period) forward rate beginning two years (4 periods) from now
${}_8f_{10}$	four-year (eight-period) forward rate beginning five years (10 periods) from now

It can be demonstrated that the formula to compute any forward rate is

$${}_tf_m = \left[\frac{(1 + z_{m+t})^{m+t}}{(1 + z_m)^m}\right]^{1/t} - 1$$

Notice that if t is equal to 1, the formula reduces to the one-period (six-month) forward rate.

To illustrate, for the spot rates shown in Table 2, suppose that an investor wants to know the two-year forward rate three years from now. In terms of the notation, t is equal to 4 and m is equal to 6. Substituting for t and m into the equation for the forward rate we have

$${}_4f_6 = \left[\frac{(1 + z_{6+4})^{6+4}}{(1 + z_6)^6}\right]^{1/4} - 1$$

This means that the following two spot rates are needed: z_6 (the three-year spot rate) and z_{10} (the five-year spot rate). From Table 3, we know

$$z_6(\text{the three-year spot rate}) = 9.9787\%/2 = 4.894\% = 0.04894$$
$$z_{10}(\text{the five-year spot rate}) = 13.623\%/2 = 6.812\% = 0.06812$$

then

$${}_4f_6 = \left[\frac{(1.06812)^{10}}{(1.04894)^6}\right]^{1/4} - 1 = 0.09755 = 9.755\%$$

Therefore, ${}_4f_6$ is equal to 9.755%, and doubling this rate gives 19.510%, the forward rate on a bond-equivalent basis.

We can verify this result. Investing \$100 for 10 periods at the semiannual spot rate of 6.812% will produce the following value:

$$\$100(1.06812)^{10} = \$193.286.$$

By investing \$100 for six periods at 4.894% and reinvesting the proceeds for four periods at the forward rate of 9.755%, it gives the same value

$$\$100(1.04894)^{6}(1.09755)^{4} = \$193.286.$$

Relationship between Spot Rates and Short-Term Forward Rates

Suppose an investor purchases a five-year, zero-coupon Treasury security for \$58.48 with a maturity value of \$100. The investor could instead buy a six-month Treasury bill and reinvest the proceeds every six months for five years. The number of dollars that will be realized will depend on the six-month forward rates. Suppose that the investor can actually reinvest the proceeds maturing every six months at the implied six-month forward rates. Let us see how many dollars would accumulate at the end of five years. The implied six-month forward rates were calculated for the yield curve given in Table 2. Letting f_t denote the six-month forward rate beginning t six-month periods from now, then the semiannual implied forward rates using the spot rates shown in Table 2 are

$$\begin{aligned} f_1 &= 0.043000 \quad f_2 = 0.050980 \quad f_3 = 0.051005 \quad f_4 = 0.051770 \\ f_5 &= 0.056945 \quad f_6 = 0.060965 \quad f_7 = 0.069310 \quad f_8 = 0.064625 \\ f_9 &= 0.062830 \end{aligned}$$

By investing the \$58.48 at the six-month spot rate of 4% (8% on a bond-equivalent basis) and reinvesting at the foregoing forward rates, the number of dollars accumulated at the end of five years would be

$$\begin{aligned} &\$58.48(1.04)(1.043)(1.05098)(1.051005)(1.05177)(1.056945) \\ &\quad (1.060965)(1.069310)(1.064625)(1.06283) = \$100 \end{aligned}$$

Therefore, we see that if the implied forward rates are realized, the \$58.48 investment will produce the same number of dollars as an investment in a five-year, zero-coupon Treasury security at the five-year spot rate. From this illustration, we can see that the five-year spot rate is related to the current six-month spot rate and the implied six-month forward rates.

In general, the relationship between a t-period spot rate, the current six-month spot rate, and the implied six-month forward rates is as follows:

$$z_t = [(1 + z_1)(1 + f_1)(1 + f_2)(1 + f_3)\cdots(1 + f_{t-1})]^{1/t} - 1$$

To illustrate how to use this equation, look at how the five-year (10-period) spot rate is related to the six-month forward rates. Substituting into the preceding equation the relevant forward rates just given and the one-period spot rate of 4% (one-half the 8% annual spot rate), we obtain

$$\begin{aligned} z_{10} &= [(1.04)(1.043)(1.05098)(1.051005)(1.05177)(1.056945) \\ &\quad (1.060965)(1.069310)(1.064625)(1.06283))]^{1/10} - 1 \\ &= 5.51\% \end{aligned}$$

Doubling 5.51% gives an annual spot rate of 11.02%, which agrees with the spot rate given in Table 2.

Forward Rate as a Hedgeable Rate

A natural question about forward rates is how well they do at predicting future interest rates. Studies have demonstrated that forward rates do not do a good job in predicting future interest rates.[4] Then, why the big deal about understanding forward rates? The reason, as we demonstrated in our illustration of how to select between two alternative investments, is that the forward rates indicate how an investor's expectations must differ from the market consensus in order to make the correct decision.

In our illustration, the six-month forward rate may not be realized. That is irrelevant. The fact is that the six-month forward rate indicated to the investor that if expectations about the six-month rate six months from now are less than 8.6%, the investor would be better off with Alternative 1.

For this reason, as well as others explained later, some market participants prefer not to talk about forward rates as being market consensus rates. Instead, they refer to forward rates as being *hedgeable rates.* For example, by buying the one-year security, the investor was able to hedge the six-month rate six months from now.

HISTORICAL SHAPES OBSERVED FOR THE TREASURY YIELD CURVE

If we plot the term structure—the yield to maturity, or the spot rate, at successive maturities against maturity—what is it likely to look like? Figure 1 shows three *generic* shapes that have appeared for the U.S. Treasury yield curve with some frequency over time. Table 4 shows five selective daily Treasury yield curves in tabular form.

Panel A of Figure 1 shows an upward-sloping yield curve; that is, yield rises steadily as maturity increases. This shape is commonly referred to as a **positively sloped yield curve.** Market participants differentiate positively sloped yield curves based on the steepness or slope of the curve. The slope is commonly measured in terms of the maturity spread, where the maturity spread is the difference between long-term and short-term yields. While there are many maturity candidates to proxy for long-term and short-term yields, we will just use the maturity spread between the six-month and 30-year yield in our example.

The first two daily yield curves shown in Table 4 are positively sloped yield curves. Notice that the three-month and six-month yields are roughly the same for both dates.

Table 4 U.S. Treasury Yield Curve for Five Selected Dates

Day	3 months	6 months	1 year	2 years	3 years	5 years	7 years	10 years	20 years	30 years	Shape
4/23/2001	3.75	3.78	3.75	3.77	4.15	4.38	4.78	5.06	5.84	5.73	Normal
4/10/1992	3.74	3.88	4.12	5.16	5.72	6.62	7.03	7.37	N/A	7.89	Steep
8/14/1981	N/A	N/A	16.71	16.91	15.88	15.34	15.04	NA	14.74	13.95	Inverted
1/3/1990	7.89	7.94	7.85	7.94	7.96	7.92	8.04	7.99	N/A	8.04	Flat
1/4/2001	5.37	5.20	4.82	4.77	4.78	4.82	5.07	5/03	5.56	5.44	Humped

Note: The data for 4/23/2001, 4/10/1992, 1/3/90, and 1/4/2001 were obtained from the daily yield curves provided by the U.S. Treasury. The data for 8/14/81 were obtained from various Treasury yield tables published by the U.S. Treasury.

[4] Eugene F. Fama, "Forward Rates as Predictors of Future Spot Rates," *Journal of Financial Economics,* 3, no. 4 (1976), pp. 361–377.

However, the steepness of the slope is different. The maturity spread between the 30-year and six-month yield (i.e., the three-month/six-month spread) was 195 basis points (5.73% to 3.78%) on 4/23/2001 and 401 basis points (7.89% to 3.88%) on 4/10/1992. The convention in the marketplace is to refer to a positively sloped yield curve as a **normal yield curve** when the spread between the six-month and 30-year yield is 300 basis points or less; when the maturity spread is more than 300 basis points, the yield curve is said to be a **steep yield curve.**

When a yield curve's maturity spread increases (or in the parlance of the market, it "widens"), the yield curve is said to *steepen*; when the maturity spread decreases (i.e., "narrows"), the yield curve is said to *flatten.*

Panel B of Figure 1 shows a downward-sloping or **inverted yield curve**, where yields are in general decline as maturity increases. There have not been many instances in the recent history of the U.S. Treasury market where the yield curve exhibited this characteristic. The most notable example is in August 1981. Table 4 shows the daily yield curve for one day in that month, August 14th. Treasury yields at the time were at a historic high. The yield on the two-year was 16.91% and declined for each subsequent maturity until it reached 13.95% for the 30-year maturity.

Finally, panel C of Figure 1 shows a **flat yield curve.** While the figure suggests that in a flat yield curve the yields are identical for each maturity, that is not what is observed in the maturity. Rather, the yields for all maturities are similar. See the yield curve on 1/3/1990. Notice the very small six-month/30-year maturity spread: 10 basis points. A variant of the flat yield is one in which the yields on short-term and long-term Treasuries are similar but the yield on intermediate-term Treasuries is lower.

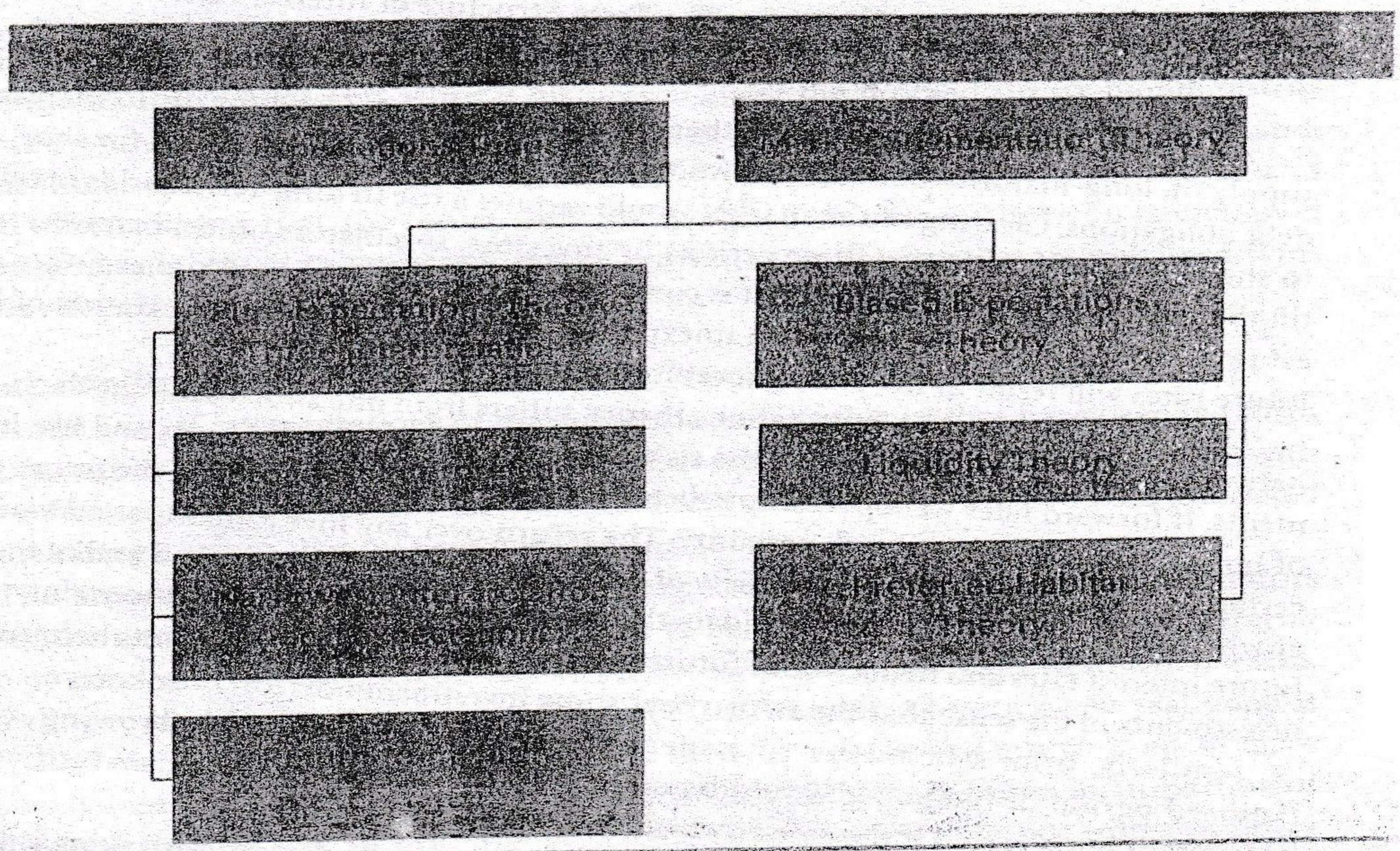
DETERMINANTS OF THE SHAPE OF THE TERM STRUCTURE

Two major theories have evolved to account for these observed shapes of the yield curve: the **expectations theory** and the **market segmentation theory.**

There are several forms of the expectations theory—the **pure expectations theory**, the **liquidity theory**, and the **preferred habitat theory.** All share a hypothesis about the behavior of short-term forward rates and also assume that the forward rates in current long-term contracts are closely related to the market's expectations about future short-term rates. These three theories differ, however, on whether or not other factors also affect forward rates, and how. The pure expectations theory postulates that no systematic factors other than expected future short-term rates affect forward rates; the liquidity theory and the preferred habitat theory assert that there are other factors. Accordingly, the last two forms of the expectations theory are sometimes referred to as **biased expectations theories.** Figure 3 depicts the relationship between these three theories.

The Pure Expectations Theory

According to the pure expectations theory, the forward rates exclusively represent the expected future rates. Thus, the entire term structure at a given time reflects the market's current expectations of the family of future short-term rates. Under this view, a rising term structure, as in panel A of Figure 1, must indicate that the market expects short-term rates to rise throughout the relevant future. Similarly, a flat term structure reflects an expectation that future short-term rates will be mostly constant, while a falling term structure must reflect an expectation that future short rates will decline steadily.



Source: *Frank J. Fabozzi, Valuation of Fixed Income Securities and Derivatives* (New Hope, PA: Frank J. Fabozzi Associates, 1995), p. 49.

We can illustrate this theory by considering how an expectation of a rising short term future rate would affect the behavior of various market participants, so as to result in a rising yield curve. Assume an initially flat term structure, and suppose that economic news subsequently leads market participants to expect interest rates to rise.

1. Those market participants interested in a long-term investment would not want to buy long-term bonds because they would expect the yield structure to rise sooner or later, resulting in a price decline for the bonds and a capital loss on the long-term bonds purchased. Instead, they would want to invest in short-term debt obligations until the rise in yield had occurred, permitting them to reinvest their funds at the higher yield.
2. Speculators expecting rising rates would anticipate a decline in the price of long-term bonds and, therefore, would want to sell any long-term bonds they own and possibly to "short sell" some they do not now own.[5] (Should interest rates rise as expected, the price of longer-term bonds will fall. Since the speculator sold these bonds short and can then purchase them at a lower price to cover the short sale, a profit will be earned.) The proceeds received from the selling of long-term debt issues the speculators now hold or the shorting of longer-term bonds will be invested in short-term debt obligations.
3. Borrowers wishing to acquire long-term funds would be pulled toward borrowing now, in the long end of the market, by the expectation that borrowing at a later time would be more expensive.

[5] Short selling means selling a security that is not owned but borrowed.

All these responses would tend either to lower the net demand for, or to increase the supply of, long-maturity bonds, and two responses would increase demand for short-term debt obligations. Clearing of the market would require a rise in long-term yields in relation to short-term yields; that is, these actions by investors, speculators, and borrowers would tilt the term structure upward until it is consistent with expectations of higher future interest rates. By analogous reasoning, an unexpected event leading to the expectation of lower future rates will result in the yield curve sloping down.

Unfortunately, the pure expectations theory suffers from one shortcoming, which, qualitatively, is quite serious. It neglects the risks inherent in investing in bonds and like instruments. If forward rates were perfect predictors of future interest rates, then the future prices of bonds would be known with certainty. The return over any investment period would be certain and independent of the maturity of the instrument initially acquired and of the time at which the investor needed to liquidate the instrument. However, with uncertainty about future interest rates and hence about future prices of bonds, these instruments become risky investments in the sense that the return over some investment horizon is unknown.

Similarly, from a borrower or issuer's perspective, the cost of borrowing for any required period of financing would be certain and independent of the maturity of the instrument initially sold if the rate at which the borrower must refinance debt in the future is known. But with uncertainty about future interest rates, the cost of borrowing is uncertain if the borrower must refinance at some time over the periods in which the funds are initially needed.

In the following section, we examine more closely the sources and types of risk that the pure expectations theory ignores.

Risks Associated with Bond Investment

There are two risks that cause uncertainty about the return over some investment horizon. The first is the uncertainty about the price of the bond at the end of the investment horizon. For example, an investor who plans to invest for five years might consider the following three investment alternatives: (1) invest in a five-year bond and hold it for five years, (2) invest in a 12-year bond and sell it at the end of five years, and (3) invest in a 30-year bond and sell it at the end of five years. The return that will be realized for the second and third alternatives is not known because the price of each long-term bond at the end of five years is not known. In the case of the 12-year bond, the price will depend on the yield on seven-year debt securities five years from now; and the price of the 30-year bond will depend on the yield on 25-year bonds five years from now. Since forward rates implicit in the current term structure for a future 12-year bond and a future 25-year bond are not perfect predictors of the actual future rates, there is uncertainty about the price for both bonds five years from now.

The risk that the price of the bond will be lower than currently expected at the end of the investment horizon is called **price risk**. An important feature of price risk is that it is greater the longer the maturity of the bond. The reason is that the longer the maturity, the greater the price volatility of a bond when yields rise. Thus, investors are exposed to price risk when they invest in a bond that will be sold prior to the bond's maturity date.

The second risk has to do with uncertainty about the rate at which the proceeds from a bond that matures prior to the maturity date can be reinvested until the maturity date. For example, an investor who plans to invest for five years might consider the following three

alternative investments: (1) invest in a five-year bond and hold it for five years; (2) invest in a six-month instrument and, when it matures, reinvest the proceeds in six-month instruments over the entire five-year investment horizon; and (3) invest in a two-year bond and, when it matures, reinvest the proceeds in a three-year bond. The risk in the second and third alternatives is that the return over the five-year investment horizon is unknown because rates at which the proceeds can be reinvested until maturity are unknown. This risk is referred to as **reinvestment risk**.

Interpretations of the Pure Expectations Theory

There are several interpretations of the pure expectations theory that have been put forth by economists. These interpretations are not exact equivalents, nor are they consistent with each other, in large part because they offer different treatments of the two risks associated with realizing a return that we have just explained.[6]

The broadest interpretation of the pure expectations theory suggests that investors expect the return for any investment horizon to be the same, regardless of the maturity strategy selected.[7] For example, consider an investor who has a five-year investment horizon. According to this theory, it makes no difference if a five-year, 12-year, or 30-year bond is purchased and held for five years because the investor expects the return from all three bonds to be the same over five years. A major criticism of this very broad interpretation of the theory is that, because of price risk associated with investing in bonds with a maturity greater than the investment horizon, the expected returns from these three very different bond investments should differ in significant ways.[8]

A second interpretation, referred to as the *local expectations* form of the pure expectations theory, suggests that the return will be the same over a short-term investment horizon starting today. For example, if an investor has a six-month investment horizon, buying a five-year, 10-year, or 20-year bond will produce the same six-month return. It has been demonstrated that the local expectations formulation, which is narrow in scope, is the only one of the interpretations of the pure expectations theory that can be sustained in equilibrium.[9]

The third and final interpretation of the pure expectations theory suggests that the return that an investor will realize by rolling over short-term bonds to some investment horizon will be the same as holding a zero-coupon bond with a maturity that is the same as that investment horizon. (A zero-coupon bond has no reinvestment risk, so future interest rates over the investment horizon do not affect the return.) This variant is called the *return-to-maturity expectations* interpretation.

For example, let's once again assume that an investor has a five-year investment horizon. By buying a five-year zero-coupon bond and holding it to maturity, the investor's return is the difference between the maturity value and the price of the bond, all divided by the price of the bond. According to return-to-maturity expectations, the same return will be realized by buying a six-month instrument and rolling it over for five years. At this time, the validity of this interpretation is subject to considerable doubt.

---

[6] These formulations are summarized by John Cox, Jonathan Ingersoll, Jr., and Stephen Ross, "A Re-examination of Traditional Hypotheses About the Term Structure of Interest Rates," *Journal of Finance* (September 1981), pp. 769–799.

[7] F. Lutz, "The Structure of Interest Rates," *Quarterly Journal of Economics* (1940–41), pp. 36–63.

[8] Cox, Ingersoll, and Ross, "A Re-examination of Traditional Hypotheses," pp. 774–775.

[9] Cox, Ingersoll, and Ross, "A Re-examination of Traditional Hypotheses," pp. 774–775.

The Liquidity Theory

We have explained that the drawback of the pure expectations theory is that it does not consider the risks associated with investing in bonds. Nonetheless, we have just shown that there is indeed risk in holding a long-term bond for one period, and that risk increases with the bond's maturity because maturity and price volatility are directly related.

Given this uncertainty, and the reasonable consideration that investors typically do not like uncertainty, some economists and financial analysts have suggested a different theory. This theory states that investors will hold longer-term maturities if they are offered a long-term rate higher than the average of expected future rates by a risk premium that is positively related to the term to maturity.[10] Put differently, the forward rates should reflect both interest rate expectations and a "liquidity" premium (really a risk premium), and the premium should be higher for longer maturities.

According to this theory, which is called the **liquidity theory of the term structure**, the implicit forward rates will not be an unbiased estimate of the market's expectations of future interest rates because they embody a liquidity premium. Thus, an upward-sloping yield curve may reflect expectations that future interest rates either (1) will rise, or (2) will be flat or even fall, but with a liquidity premium increasing fast enough with maturity so as to produce an upward-sloping yield curve.

The Preferred Habitat Theory

Another theory, known as the preferred habitat theory, also adopts the view that the term structure reflects the expectation of the future path of interest rates as well as a risk premium. However, the habitat theory rejects the assertion that the risk premium must rise uniformly with maturity.[11] Proponents of the habitat theory say that the latter conclusion could be accepted if all investors intend to liquidate their investment at the first possible date, while all borrowers are eager to borrow long, but that this is an assumption that cannot be rejected for a number of reasons.

In the first place, it is obvious that many investors wish to carry resources forward for appreciable periods of time—to buy a house, for example, or to provide for retirement. Such investors are concerned with the amount available at the appropriate time, and not the path by which that goal is reached. Hence, risk aversion dictates that they should prefer an instrument with a maturity matching the period for which they wish to invest over shorter-term investment vehicles. If these investors buy a shorter instrument, they will bear reinvestment risk—the risk of a fall in the interest rates available for reinvesting proceeds of the shorter instrument. Investors can avoid that risk only by locking in the current long rate through a long-term contract. Similarly, if they buy an instrument with maturity longer than the time they wish to invest for, they will bear the risk of a loss in the price of the asset (price risk) when liquidating it before its maturity, because of a rise in interest rates. Entirely analogous considerations apply to borrowers; prudence and safety call for borrowing for a maturity by matching the length of time for which funds are required.

---

[10] John R. Hicks, *Value and Capital*, 2nd ed. (London: Oxford University Press, 1946), pp. 141–145.

[11] Franco Modigliani and Richard Sutch, "Innovations in Interest Rate Policy," *American Economic Review* (May 1966), pp. 178–197.

In the second place, a lot of the demand for and supply of securities these days comes from financial intermediaries, which have liabilities with specified maturities. These institutions seek to match as closely as possible the maturity of their liabilities with the cash flow of a portfolio of assets. In constructing such a portfolio, a financial institution will restrict its investments to certain maturity sectors.

To illustrate this preference for maturity sectors, consider a life insurance company that has issued a five-year guaranteed investment contract. The insurance company will not want to invest in six-month instruments because of the associated reinvestment risk. As another example, assume a thrift has borrowed funds at a fixed rate for one year with the proceeds from the issuance of a one-year certificate of deposit. The thrift is exposed to price (or interest rate) risk if the borrowed funds are invested in a bond with 20 years to maturity. Clearly, then, either of these institutions faces some kind of risk if it invests outside its preferred maturity sector.

The preferred habitat theory asserts that, to the extent that the demand and supply of funds in a given maturity range does not match, some lenders and borrowers will be induced to shift to maturities showing the opposite imbalances. However, they will need to be compensated by an appropriate risk premium whose magnitude will reflect the extent of aversion to either price or reinvestment risk.

Thus, this theory proposes that the shape of the yield curve is determined by both expectations of future interest rates and a risk premium, positive or negative, to induce market participants to shift out of their preferred habitat. Clearly, according to this theory, yield curves sloping up, down, flat, or humped are all possible.

Market Segmentation Theory

The market segmentation theory also recognizes that investors have preferred habitats dictated by saving and investment flows. This theory also proposes that the major reason for the shape of the yield curve lies in asset/liability management constraints (either regulatory or self-imposed) and/or creditors (borrowers) restricting their lending (financing) to specific maturity sectors.[12] However, the market segmentation theory differs from the preferred habitat theory in that it assumes that neither investors nor borrowers are willing to shift from one maturity sector to another to take advantage of opportunities arising from differences between expectations and forward rates.

Thus, for the segmentation theory, the shape of the yield curve is determined by supply of and demand for securities within each maturity sector. This formulation seems untenable because it presupposes the prevalence of absolute risk aversion, while the evidence does not support that proposition. Thus, market participants must be expected to shift away from their habitat when there are sufficiently large discrepancies between market and expected rates. This potential shifting ensures that the differences between market and expected rates will not grow too large, and this consideration leads back to the preferred habitat theory.

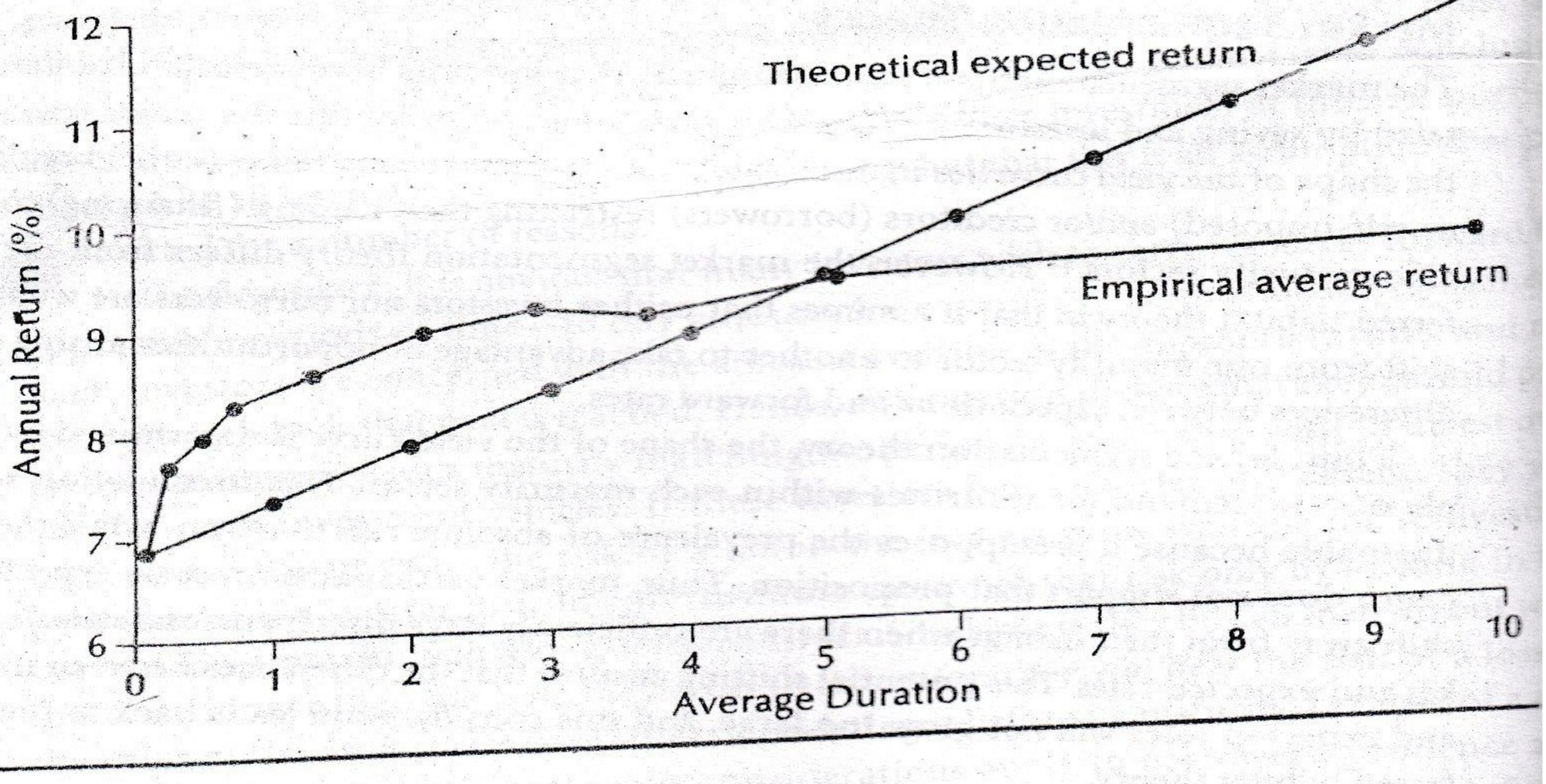
THE MAIN INFLUENCES ON THE SHAPE OF THE YIELD CURVE

Comprehensive research on the main influences of the shape of the Treasury yield curve was done by Antti Ilmanen in a series of papers.[13] He finds that there are three main influences. The first is the market's expectations of future rate changes as explained earlier when we discussed the pure expectations theory. The other two are bond risk premiums and convexity bias.

Bond risk premiums are the expected return differentials across Treasury securities of different maturities. As explained in the previous section, there are theories of the term structure of interest rates that hypothesize why expected returns will vary by maturity. However, the theories disagree with respect to whether the risk premium is positive or negative. For example, the liquidity theory of the term structure would argue that the risk premium should increase with maturity; the market segmentation theory says that the bond risk premium can be positive or negative.

Ilmanen investigated the effect of the behavior of the bond risk premium using historical average *returns* on U.S. Treasury securities. Figure 4 shows the empirical average return curve as a function of average duration (not maturity) for the period 1972 to 2001. (Duration is a measure of the price sensitivity of a bond to changes in interest rates.) Also shown in the figure is the theoretical expected return curve based on expectations only (the first influence listed previously). Notice that this curve is linear (i.e., it increases linearly with duration). In contrast, notice that the empirical evidence suggests that the bond risk

Figure 4 Theoretical [illegible]



[13] The research first appeared as a report published by Salomon Brothers. Different parts of the research report were then published. See Antti Ilmanen, "Market's Rate Expectations and Forward Rates," *Journal of Fixed Income* (September 1996), pp. 8–22; Antti Ilmanen, "Does Duration Extension Enhance Long-Term Expected Returns?" *Journal of Fixed Income* (September 1996), pp. 23–36; Antti Ilmanen, "Convexity Bias in the Yield Curve," Chapter 3 in Narasimgan Jegadeesh and Bruce Tuckman (eds.), *Advanced Fixed-Income Valuation Tools* (New York: Wiley, 2000); and Antti Ilmanen, "Overview of Forward Rate Analysis," Chapter 8 in Frank J. Fabozzi (ed.), *The Handbook of Fixed Income Securities* (New York: McGraw Hill, 2005).

premiums are not linear in duration. Instead, the empirical evidence suggests that at the front end of the yield curve (i.e., up to a duration of 3), bond risk premiums increase steeply with duration. However, after a duration of 3, the bond risk premiums increase slowly. Ilmanen suggests that the shape shown in Figure 4 "may reflect the demand for long-term bonds from pension funds and other long-duration liability holders."[14] Pension funds seek long-term bonds to lock in rates to satisfy their liability obligation.

Now let's look to the convexity bias influence, an influence that Ilmanen argues is the least well known of the three influences. When interest rates change by a large number of basis points, a Treasury security's price change will not be the same for an increase and decrease in interest rates. More specifically, the price appreciation when interest rates fall will be greater than the price decline when interest rates rise by the same number of basis points. For example, if interest rates decline by 100 basis points, the price of a Treasury security might appreciate by 20%, but if interest rates increase by 100 basis points, the price of the same Treasury security might decline by only 15%. This attractive property of a bond is due to the shape of the relationship between price and yield and is referred to as the bond's convexity. The longer the maturity, the more convexity the security has. That is, longer-term Treasury securities have a more attractive feature due to convexity than shorter-term Treasury securities. As a result, investors are willing to pay more for longer-term Treasury securities and therefore accept lower returns. This influence on the shape of the Treasury yield curve is what is referred to as the *convexity bias*.

Key Points That You Should Understand Before Proceeding

1. The different shapes that have been observed for the yield curve.
2. The common hypothesis about the behavior of short-term forward rates shared by the various forms of the expectations theory.
3. The implications of the shape of the yield curve for future interest rates based on the pure expectations theory.
4. The two types of risks associated with investing in bonds and how these two risks affect the pure expectations theory.
5. How the two biased expectations theories differ from the pure expectations theory.
6. The three main influences on the shape of the Treasury yield curve are (1) the market's expectations of future rate changes, (2) bond risk premiums, and (3) convexity bias.

SUMMARY

The relationship between yield and maturity is referred to as the term structure of interest rates. The graphical depiction of the relationship between the yield on bonds of the same credit quality but different maturities is known as the yield curve. Since historically the yield on Treasury securities is the base rate off which a nongovernment bond's yield often is benchmarked, the most commonly constructed yield curve is the Treasury yield curve.

There is a problem with using the Treasury yield curve to determine the one yield at which to discount all the cash payments of any bond. Each cash flow within a bond's total

14 Ilmanen, "Overview of Forward Rate Analysis," p. 167.

pattern of cash flows should be discounted at a unique interest rate that is applicable to the time period when the cash flow is to be received. Since any bond can be viewed as a package of zero-coupon instruments, its value should equal the value of all the component zero-coupon instruments. The rate on a zero-coupon bond is called the spot rate. The theoretical spot rate curve for Treasury securities can be estimated from the Treasury yield curve using a methodology known as bootstrapping.

Under certain assumptions, the market's expectation of future interest rates can be extrapolated from the theoretical Treasury spot rate curve. The resulting forward rate is called the implicit forward rate. The spot rate is related to the current six-month spot rate and the implicit six-month forward rates. A knowledge of the forward rates implicit in the current long-term rate is relevant in formulating both investment strategies and borrowing policies.

Several theories have been proposed about the determination of the term structure. The pure expectations theory hypothesizes that the one-period forward rates simply represent the market's expectations of future actual rates. Thus, the long-term spot rate would itself be explained fully by the market expectations of future short rates. The term structure might then be rising, falling, or flat, according to whether the market expects rising, falling, or unchanged short-term rates. This formulation fails to recognize the risks associated with investing in bonds—price risk and reinvestment risk—when investors buy bonds whose maturity is different from the time for which they plan to hold the bond.

The fact that there is price risk in investing in long-term bonds, and that it seems to increase with maturity, has given rise to an alternative liquidity theory of the term structure. According to this theory, forward rates are the sum of expected future rates and a risk premium that increases for more and more distant future rates and, hence, rises with the maturity of a bond. This formulation has shortcomings because it presupposes that all lenders want to lend short and all borrowers want to borrow long. If so, long borrowers would have to offer lenders a premium, rising with maturity, to accept the risk of going long. But in reality both lenders and borrowers have quite varied maturity preferences. Each market participant can eliminate risk, not by borrowing or lending short, but by lending (or borrowing) for a period coinciding with their preferred habitat. But, at the same time, agents would presumably be willing to depart from their preferred habitat by the inducement of a risk premium.

Accordingly, the third version of the expectations theory—namely, the preferred habitat theory—suggests, like the liquidity theory, that forward rates are the sum of a component reflecting expected future rates and a risk premium. However, the premium will not rise continuously with maturity but will materialize in any maturity neighborhood where supply exceeds demand. A negative premium or discount would be expected if supply exceeds demand.

One final theory explains the shape of the term structure by the concept of market segmentation. In common with the preferred habitat theory, it recognizes that participants in the bond market have maturity preferences. However, it postulates that these preferences are absolute and cannot be overcome by the expectation of a higher return from a different maturity, no matter how large. Each maturity is therefore a separate market, and the interest rate in every such market is determined by the given demand and supply. Thus, the interest rate at any maturity is totally unrelated to expectations of future rates. This formulation is of doubtful use because it implies highly irrational, implausible, and counterfactual behavior.

Empirical evidence suggests that the three main influences on the shape of the Treasury yield curve are (1) the market's expectations of future rate changes, (2) bond risk premiums, and (3) convexity bias.

KEY TERMS

- Biased expectations theories
- Bootstrapping
- Convexity bias
- Expectations theory
- Flat yield curve
- Forward rate (implied forward rate)
- Inverted yield curve
- Liquidity theory
- Liquidity theory of the term structure
- Market segmentation theory
- Normal yield curve
- Positively sloped yield curve
- Preferred habitat theory
- Price risk
- Pure expectations theory
- Reinvestment risk
- Steep yield curve
- Spot rate
- Spot rate curve
- Term structure of interest rates
- Yield curve
- Zero-coupon instrument

QUESTIONS

1. a. What is a yield curve?
   b. Historically, why has the Treasury yield curve been the one that is most closely watched by market participants?
2. What is meant by a spot rate?
3. Explain why it is inappropriate to use one yield to discount all the cash flows of a financial asset.
4. Explain why a financial asset can be viewed as a package of zero-coupon instruments.
5. Why is it important for lenders and borrowers to have a knowledge of forward rates?
6. How are spot rates related to forward rates?
7. You are a financial consultant. At various times, you have heard the following comments on interest rates from one of your clients. How would you respond to each comment?
   a. "The yield curve is upward sloping today. This suggests that the market consensus is that interest rates are expected to increase in the future."
   b. "I can't make any sense out of today's term structure. For short-term yields (up to three years), the spot rates increase with maturity; for maturities greater than three years but less than eight years, the spot rates decline with maturity; and for maturities greater than eight years the spot rates are virtually the same for each maturity. There is simply no theory that explains a term structure with this shape."
   c. "When I want to determine the market's consensus of future interest rates, I calculate the implicit forward rates."
8. You observe the Treasury yield curve below (all yields are shown on a bond-equivalent basis):

Year	Yield to Maturity	Spot Rate
0.5	5.25%	5.25%
1.0	5.50	5.50
1.5	5.75	5.76
2.0	6.00	?
2.5	6.25	?
3.0	6.50	?
3.5	6.75	?
4.0	7.00	?
4.5	7.25	?
5.0	7.50	?
5.5	7.75	7.97
6.0	8.00	8.27
6.5	8.25	8.59
7.0	8.50	8.92
7.5	8.75	9.25
8.0	9.00	9.61
8.5	9.25	9.97
9.0	9.50	10.36
9.5	9.75	10.77
10.0	10.00	11.20

All the securities maturing from 1.5 years on are selling at par. The 0.5-year and one-year securities are zero-coupon instruments.

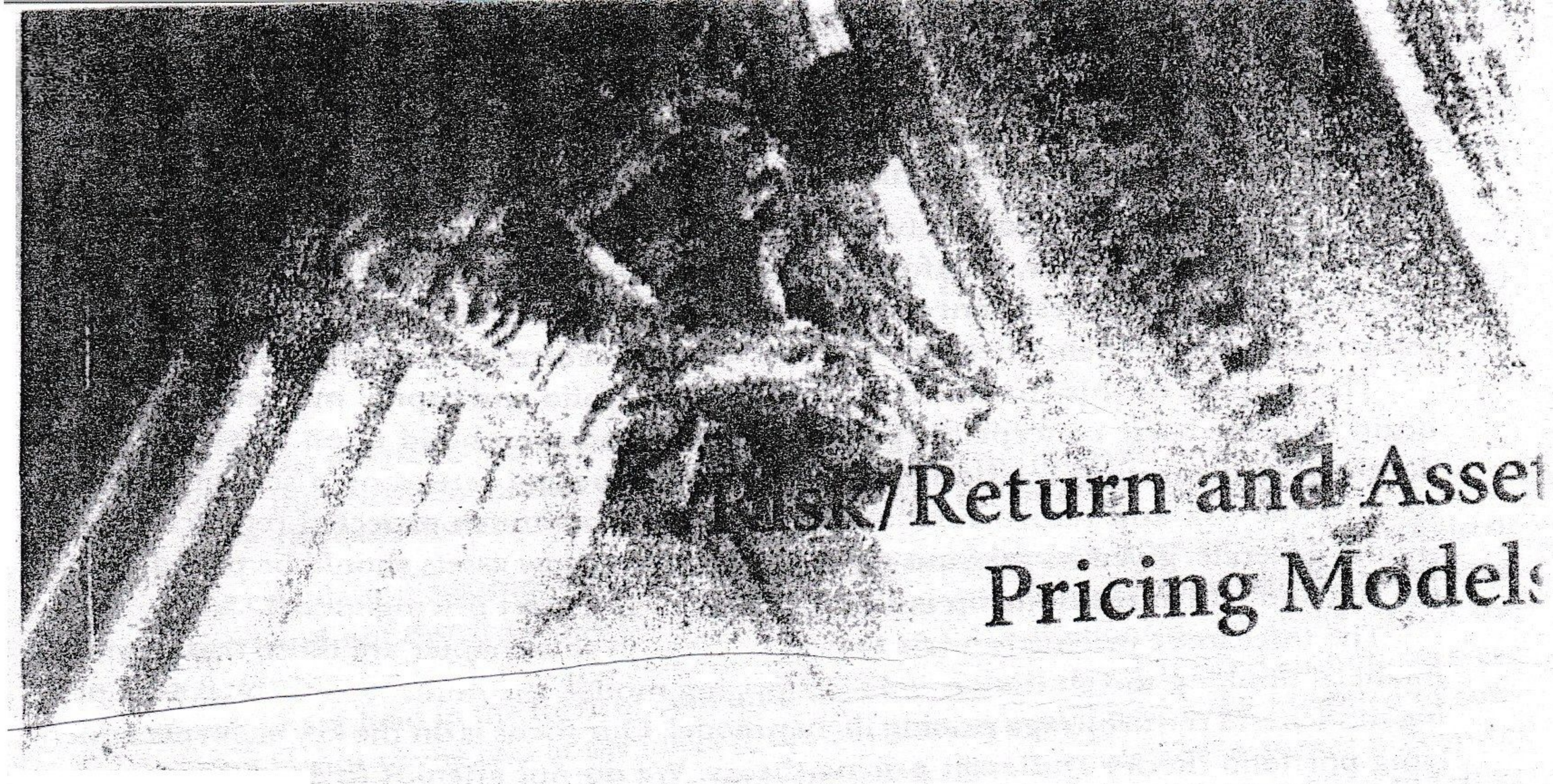
a. Calculate the missing spot rates.
b. What should the price of the six-year Treasury security be?
c. What is the implicit six-month forward rate starting in the sixth year?

9. You observe the following Treasury yield curve (all yields are shown on a bond-equivalent basis):

Year	Yield to Maturity	Spot Rate
0.5	10.00%	10.00%
1.0	9.75	9.75
1.5	9.50	9.48
2.0	9.25	9.22
2.5	9.00	8.95
3.0	8.75	8.68
3.5	8.50	8.41
4.0	8.25	8.14
4.5	8.00	7.86
5.0	7.75	7.58
5.5	7.50	7.30
6.0	7.25	7.02
6.5	7.00	6.74
7.0	6.75	6.46
7.5	6.50	6.18
8.0	6.25	5.90
8.5	6.00	5.62
9.0	5.75	5.35
9.5	5.50	?
10.0	5.25	?

All the securities maturing from 1.5 years on are selling at par. The 0.5-year and one-year securities are zero-coupon instruments.
a. Calculate the missing spot rates.
b. What should the price of the four-year Treasury security be?

10. Using the theoretical spot rates in Table 2, calculate the theoretical value of a 7%, six-year Treasury bond.
11. a. Using the theoretical spot rates in Table 2, calculate the two-year forward rate four years from now.
b. Verify the answer by assuming an investment of $100 is invested for six years.
12. Explain the role that forward rates play in making investment decisions.
13. "Forward rates are poor predictors of the actual future rates that are realized. Consequently, they are of little value to an investor." Explain why you agree or disagree with this statement.
14. An investor is considering two alternative investments. The first alternative is to invest in an instrument that matures in two years. The second alternative is to invest in an instrument that matures in one year and at the end of one year, reinvest the proceeds in a one-year instrument. The investor believes that one-year interest rates one year from now will be higher than they are today and, therefore, is leaning in favor of the second alternative. What would you recommend to this investor?
15. a. What is the difference between a normal yield curve and steep yield curve?
b. What is meant by a humped yield curve?
16. What is the common hypothesis about the behavior of short-term forward rates shared by the various forms of the expectations theory?
17. What are the types of risks associated with investing in bonds and how do these two risks affect the pure expectations theory?
18. Give three interpretations of the pure expectations theory.
19. What are the two biased expectations theories about the term structure of interest rates?
20. What are the underlying hypotheses of the two biased expectations theories of interest rates?
21. a. "Empirical evidence suggests that with respect to bond risk premiums that influence the shape of the Treasury yield curve, there is a linear relationship between Treasury average returns and duration." Explain whether you agree or disagree with this statement. If you disagree, explain the type of relationship that has been observed.
b. What is meant by the "convexity bias" influence on the shape of the Treasury yield curve?



Risk, Return and Asset Pricing Models

LEARNING OBJECTIVES

After reading this chapter, you will understand

- the fundamental principles of portfolio theory
- how to calculate the historical single-period investment return for a security or portfolio of securities
- how to calculate the expected return and variability of expected return of a portfolio
- the components of a portfolio's total risk: systematic risk and unsystematic risk
- what the beta of a stock measures
- why diversification eliminates unsystematic risk
- the capital asset pricing model, the relevant measure of risk in this model, and the limitations of the model
- the development of the multifactor capital asset pricing model
- the empirical difficulties of testing the capital asset pricing model
- the fundamental principles underlying the arbitrage pricing theory model
- the empirical difficulties of testing the arbitrage pricing theory model
- the importance of the return distribution of financial assets in finance theory
- what behavioral finance is and how it differs from standard finance theory

Portfolio theory deals with the selection of optimal portfolios by rational risk-aver investors—that is, by investors who attempt to maximize their expected portfolio retur consistent with individually acceptable levels of portfolio risk. Capital markets theory dea with the implications for security prices of the decisions made by these investors—that is, t

From Chapter 12 of *Foundations of Financial Markets and Institutions*, 4/e. Frank J. Fabozzi. Franco Modigliani. Frank J. Jone
Copyright © 2010 by Pearson Prentice Hall. All rights reserved.

relationship that should exist between security returns and risk if investors behave in this optimal fashion. Together, portfolio and capital markets theories provide a framework to specify and measure investment risk and to develop relationships between expected security return and risk (and hence between risk and required return on investment).

The purpose of this chapter is to introduce portfolio and capital markets theories. We begin with the basic concepts of portfolio theory and then build upon these concepts to develop the theoretical relationship between the expected return on a security and risk. Because the risk and return relationship indicates how much expected return a security should generate, given its relevant risks, it also tells us how assets should be priced. Hence, the risk and return relationship is also referred to as an asset pricing model.

The three asset pricing models that we present in this chapter are those that dominate financial thinking today: the capital asset pricing model, the multifactor capital asset pricing model, and the arbitrage pricing theory model. Our focus is on the key elements underlying portfolio theory and asset pricing theory. We do not attempt to provide a rigorous mathematical presentation of these theories.

Although the theories explained in this chapter are the cornerstone of much of finance, they have been under constant attack. The first attack has been on whether current financial theory is telling us the proper way to measure risk. The second is whether the way investors behave as assumed by standard finance theory is correct because it fails to take into account the behavior of investors in making investment decisions as suggested by the literature in the field of psychology. This has led to the development of a branch in finance called behavioral finance, which draws from the work of several well-known psychologists, one of whom, Daniel Kahneman, was the corecipient of the Alfred Nobel Memorial Prize in Economic Sciences in 2002. We provide a brief description of this behavioral approach to finance at the end of this chapter.

PORTFOLIO THEORY

In designing a portfolio, investors seek to maximize the expected return from their investment, given some level of risk they are willing to accept.[1] Portfolios that satisfy this requirement are called **efficient (or optimal) portfolios.**[2] To construct an efficient portfolio, it is necessary to understand what is meant by **expected return** and risk. The latter concept, risk, could mean any one of the many types of risk. We shall be more specific about its meaning as we proceed in the development of portfolio theory. We begin our exploration of portfolio theory with the concept of investment return.

Investment Return

The return on an investor's portfolio during a given interval is equal to the change in value of the portfolio plus any distributions received from the portfolio, expressed as a fraction of the initial portfolio value. It is important that any capital or income distributions made to the investor be included, or the measure of return will be deficient. Another way to look at

---

[1] Alternatively stated, investors seek to minimize the risk to which they are exposed, given some target expected return.

[2] The theoretical framework for selecting efficient portfolios was developed in Harry M. Markowitz, "Portfolio Selection," *Journal of Finance* (March 1952), pp. 71–91, and *Portfolio Selection: Efficient Diversification of Investments* (New York: John Wiley & Sons, Inc., 1959).

return is as the amount (expressed as a fraction of the initial portfolio value) that can be withdrawn at the end of the interval while maintaining the principal intact. The return on the investor's portfolio, designated R_p, is given by

$$R_p = \frac{V_1 - V_0 + D_1}{V_0} \tag{1}$$

where

V_1 = the portfolio market value at the end of the interval
V_0 = the portfolio market value at the beginning of the interval
D_1 = the cash distributions to the investor during the interval

The calculation assumes that any interest or dividend income received on the portfolio of securities and not distributed to the investor is reinvested in the portfolio (and thus reflected in V_1). Furthermore, the calculation assumes that any distributions occur at the end of the interval, or are held in the form of cash until the end of the interval. If the distributions were reinvested prior to the end of the interval, the calculation would have to be modified to consider the gains or losses on the amount reinvested. The formula also assumes no capital inflows during the interval. Otherwise, the calculation would have to be modified to reflect the increased investment base. Capital inflows at the end of the interval (or held in cash until the end), however, can be treated as just the reverse of distributions in the return calculation.

Thus, given the beginning and ending portfolio values, plus any contributions from or distributions to the investor (assumed to occur at the end of an interval), equation (1) lets us compute the investor's return. For example, if the XYZ pension fund had a market value of $100,000 at the end of June, a distribution of $5,000 made at the end of July, and an end-of-July market value of $103,000, the return for the month would be 8%:

$$R_p = \frac{103{,}000 - 100{,}000 + 5{,}000}{100{,}000} = 0.08$$

In principle, this sort of calculation of returns could be carried out for any interval of time, say, for one month or 10 years. Yet, there are several problems with this approach. First, it is apparent that a calculation made over a long period of time, say, more than a few months, would not be very reliable because of the underlying assumption that all cash payments and inflows are made and received at the end of the period. Clearly, if two investments have the same return as calculated from the formula above, but one investment makes a cash payment early and the other late, the one with early payment will be understated. Second, we cannot rely on the formula to compare return on a one-month investment with that on a 10-year return portfolio. For purposes of comparison, the return must be expressed per unit of time—say, per year.

In practice, we handle these two problems by first computing the return over a reasonably short unit of time, perhaps a quarter of a year or less. The return over the relevant horizon, consisting of several unit periods, is computed by averaging the return over the unit intervals.[3]

---

[3] There are three generally used methods of averaging: (1) the arithmetic average return, (2) the time-weighted rate of return (also referred to as the geometric rate of return), and (3) the dollar-weighted return. The averaging yields a measure of return per unit of time period. The measure can be converted to an annual or other period yield by standard procedures. For an explanation of these three methods of averaging, see Franco Modigliani and Gerald A. Pogue, "Risk, Return, and CAPM: Concepts and Evidence," Chapter 37 in Summer N. Levine (ed.), *The Financial Analysts Handbook* (Homewood, IL: Dow Jones-Irwin, 1988).

Portfolio Risk

The definition of investment risk leads us into less explored territory. Not everyone agrees on how to define risk, let alone measure it. Nevertheless, there are some attributes of risk that are reasonably well accepted.

An investor holding a portfolio of Treasury bills until the maturity date faces no uncertainty about monetary outcome. The value of the portfolio at maturity of the securities will be identical with the predicted value; the investor bears no monetary risk. In the case of a portfolio composed of common stocks, however, it will be impossible to predict the value of the portfolio at any future date. The best an investor can do is to make a best-guess or most-likely estimate, qualified by statements about the range and likelihood of other values. In this case, the investor does bear risk.

One measure of risk is the extent to which possible future portfolio values are likely to diverge from the expected or predicted value. More specifically, risk for most investors is related to the chance that future portfolio values will be less than expected. That is, if the investor's portfolio has a current value of $100,000, and an expected value of $110,000 at the end of the next year, what matters is the probability of values less than $110,000.

Before proceeding to the quantification of risk, it is convenient to shift our attention from the terminal value of the portfolio to the portfolio rate of return, R_p, because the increase in portfolio value is related directly to R_p.

Expected Portfolio Return

A particularly useful way to quantify the uncertainty about the portfolio return is to specify the probability associated with each of the possible future returns. Assume, for example, that an investor has identified five possible outcomes for the portfolio return during the next year. Associated with each return is a subjectively determined probability, or relative chance of occurrence. The five possible outcomes are

Outcome	Possible Return	Subjective Probability
1	50%	0.1
2	30	0.2
3	10	0.4
4	−10	0.2
5	−30	0.1

Note that the probabilities sum to 1 so that the actual portfolio return is confined to one of the five possible values. Given this **probability distribution**, we can measure the expected return and risk for the portfolio.

The expected return is simply the weighted average of possible outcomes, where the weights are the relative chances of occurrence. In general, the expected return on the portfolio, denoted $E(R_p)$, is given by

$$E(R_p) = P_1R_1 + P_2R_2 + \cdots + R_nP_n \tag{2}$$

or

$$E(R_p) = \sum_{j=1}^{n} P_jR_j$$

where the R_js are the possible returns, the P_js the associated probabilities, and n is the number of possible outcomes.

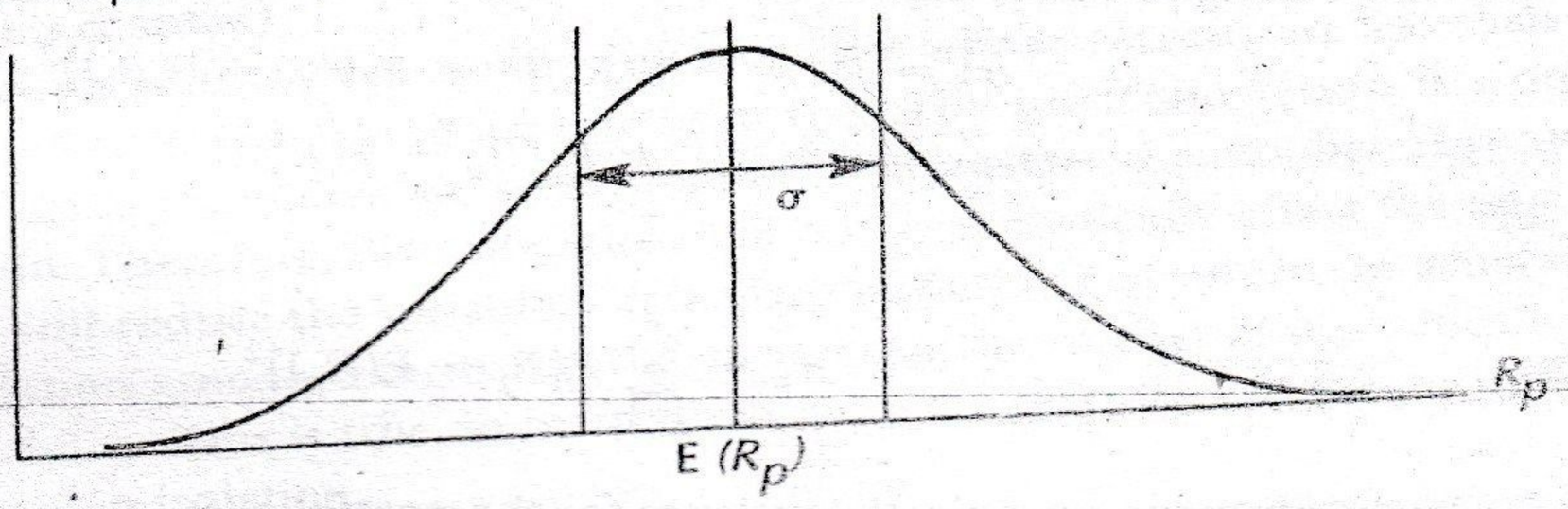
The expected return of the portfolio in our illustration is:

$$\begin{aligned} E(R_p) &= 0.1(50.0) + 0.2(30.0) + 0.4(10.0) + 0.2(-10.0) + 0.1(-30.0) \\ &= 10\% \end{aligned}$$

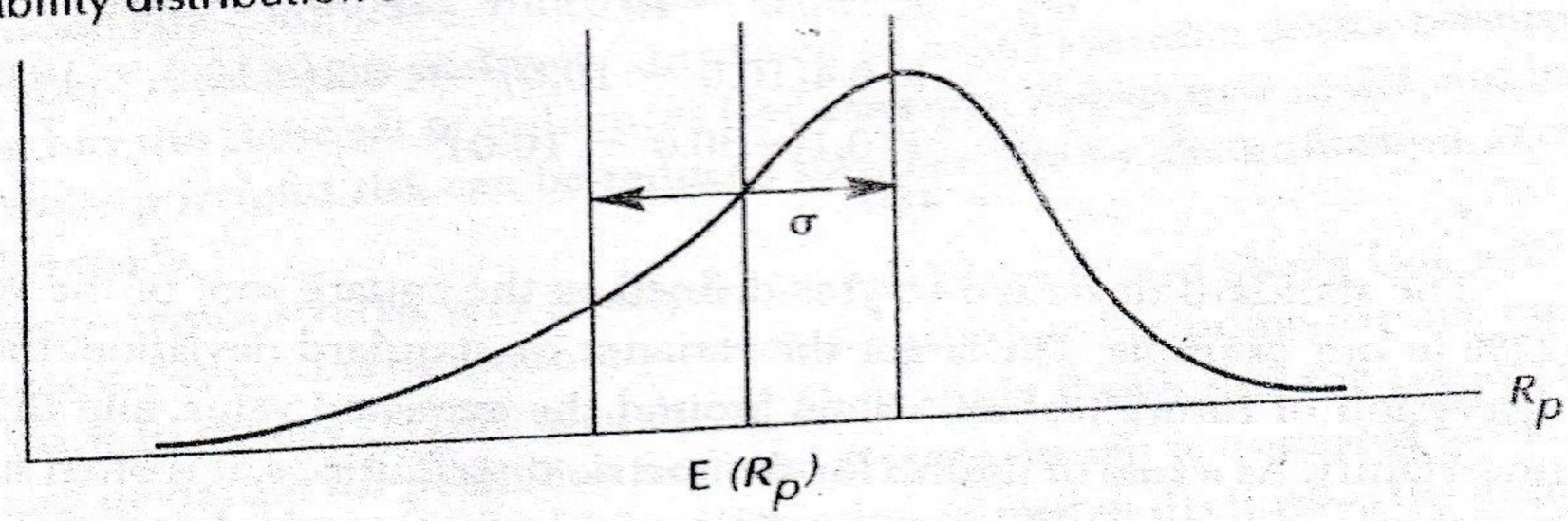
Variability of Expected Return

If risk is defined as the chance of achieving returns lower than expected, it would seem logical to measure risk by the dispersion of the possible returns below the expected value. Risk measures based on below-the-mean variability are difficult to work with, however, and moreover are unnecessary as long as the distribution of future return is reasonably symmetric about the expected value. Figure 1 shows three probability distributions: the first

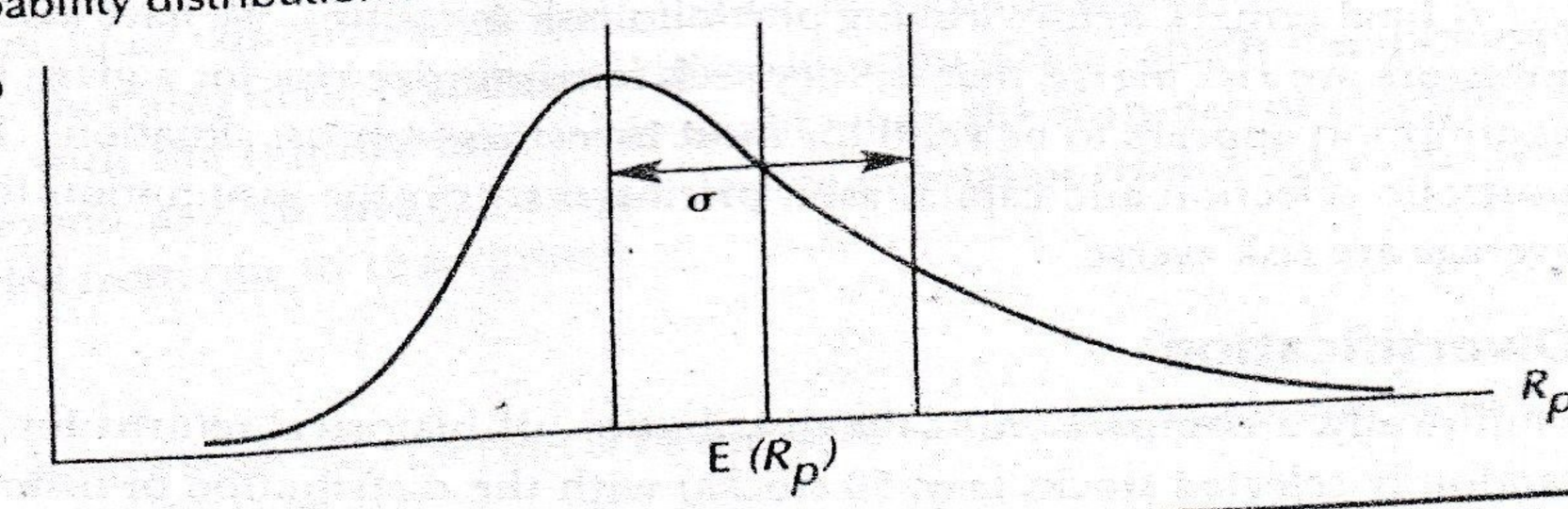
Symmetric probability distribution



Probability distribution skewed to left



Probability distribution skewed to right



symmetric, the second skewed to the left, and the third skewed to the right. For a **symmetrical distribution**, the dispersion of returns on one side of the expected return is the same as the dispersion on the other side of the expected return.

If the probability distribution is symmetric, measures of the total variability of return will be twice as large as measures of the portfolio's variability below the expected return. Thus, if total variability is used as a risk surrogate, the risk ranking for a group of portfolios will be the same as when variability below the expected return is used. It is for this reason that total variability of returns has been used so widely as a surrogate for risk. While much of financial theory has been developed assuming the distribution is symmetric, at the end of this chapter, we discuss this assumption, its relevance, and its implications.

It now remains to choose a specific measure of total variability of returns. The most commonly used measures are the **variance** and **standard deviation** of returns.

The variance of return is a weighted sum of the squared deviations from the expected return. Squaring the deviations ensures that deviations above and below the expected value contribute equally to the measure of variability regardless of sign. The variance for the portfolio, designated σ_p^2, is given by

$$\sigma_p^2 = P_1[R_1 - \mathrm{E}(R_p)]^2 + P_2[R_2 - \mathrm{E}(R_p)]^2 + \cdots + P_n[R_n - \mathrm{E}(R_p)]^2$$

or

$$\sigma_p^2 = \sum_{j=1}^{n} P_j[R_j - \mathrm{E}(R_p)]^2 \tag{3}$$

In the previous example, the variance for the portfolio is

$$\begin{aligned}\sigma_p^2 &= 0.1(50.0 - 10.0)^2 + 0.2(30.0 - 10.0)^2 \\ &\quad + 0.4(10.0 - 10.0)^2 + 0.2(-10.0 - 10.0)^2 \\ &\quad + 0.1(-30.0 - 10.0)^2 \\ &= 480\end{aligned}$$

The standard deviation (σ_p) is defined as the square root of the variance. It is equal to 22% in our example. The larger the variance or standard deviation, the greater the possible dispersion of future realized values around the expected value, and the larger the investor's uncertainty. As a rule of thumb for symmetric distributions, it is often suggested that roughly two-thirds of the possible returns will lie within one standard deviation on either side of the expected value, and that 95% will be within two standard deviations of the expected value.

A final remark before leaving portfolio risk measures. We have assumed implicitly that investors are *risk averse*; that is, they seek to minimize risk for a given level of return. This assumption appears to be valid for most investors in most situations. The entire theory of portfolio selection and capital asset pricing rests on the assumption that investors on the average are risk averse.

Diversification

Empirically, a comparison of the distribution of historical returns for a large portfolio of randomly selected stocks (say, 50 stocks) with the distribution of historical returns for an individual stock in the portfolio has indicated a curious relationship. It is not uncommon

to find that (1) the standard deviation of return for the individual stocks in the portfolio is considerably larger than that of the portfolio, and (2) the average return of an individual stock is less than the portfolio return. Is the market so imperfect that it tends to reward substantially higher risk with lower stock return?

Not so. The answer lies in the fact that not all of an individual stock's risk is relevant. Much of the **total risk** (which equals standard deviation of return) is **diversifiable**. That is, if an investment in an individual stock is combined with other securities, a portion of the variation in its returns could be smoothed or canceled by complementary variation in the other securities. The same portfolio diversification effect accounts for the low standard deviation of return for a large stock portfolio. In fact, the portfolio standard deviation is lower than that of the typical security in the portfolio. Much of the total risk of the component securities has been eliminated by diversification. As long as much of the total risk can be eliminated simply by holding a stock in a portfolio, there is no economic requirement for the anticipated return to be in line with the total risk. Instead, we should expect realized returns to be related to that portion of security risk that cannot be eliminated by portfolio combination—so-called **systematic risk**. (We will have more to say on risk/return relationships later in this chapter.)

Diversification results from combining securities whose returns are less than perfectly correlated in order to reduce portfolio risk. The portfolio return is simply a weighted average of the individual security returns, no matter the number of securities in the portfolio. Therefore, diversification will not systematically affect the portfolio return, but it will reduce the variability (standard deviation) of return. In general, the less the **correlation among security returns**, the greater the impact of diversification on reducing variability. This is true no matter how risky the securities of the portfolio are when considered in isolation.

Theoretically, if we could find sufficient securities with uncorrelated returns, we could **eliminate portfolio risk completely**. Unfortunately, this situation is not typical in real financial markets, where returns are positively correlated to a considerable degree because they tend to respond to the same set of influences (for example, to business cycles and interest rates). Thus, while portfolio risk can be reduced substantially by diversification, it cannot be eliminated entirely.

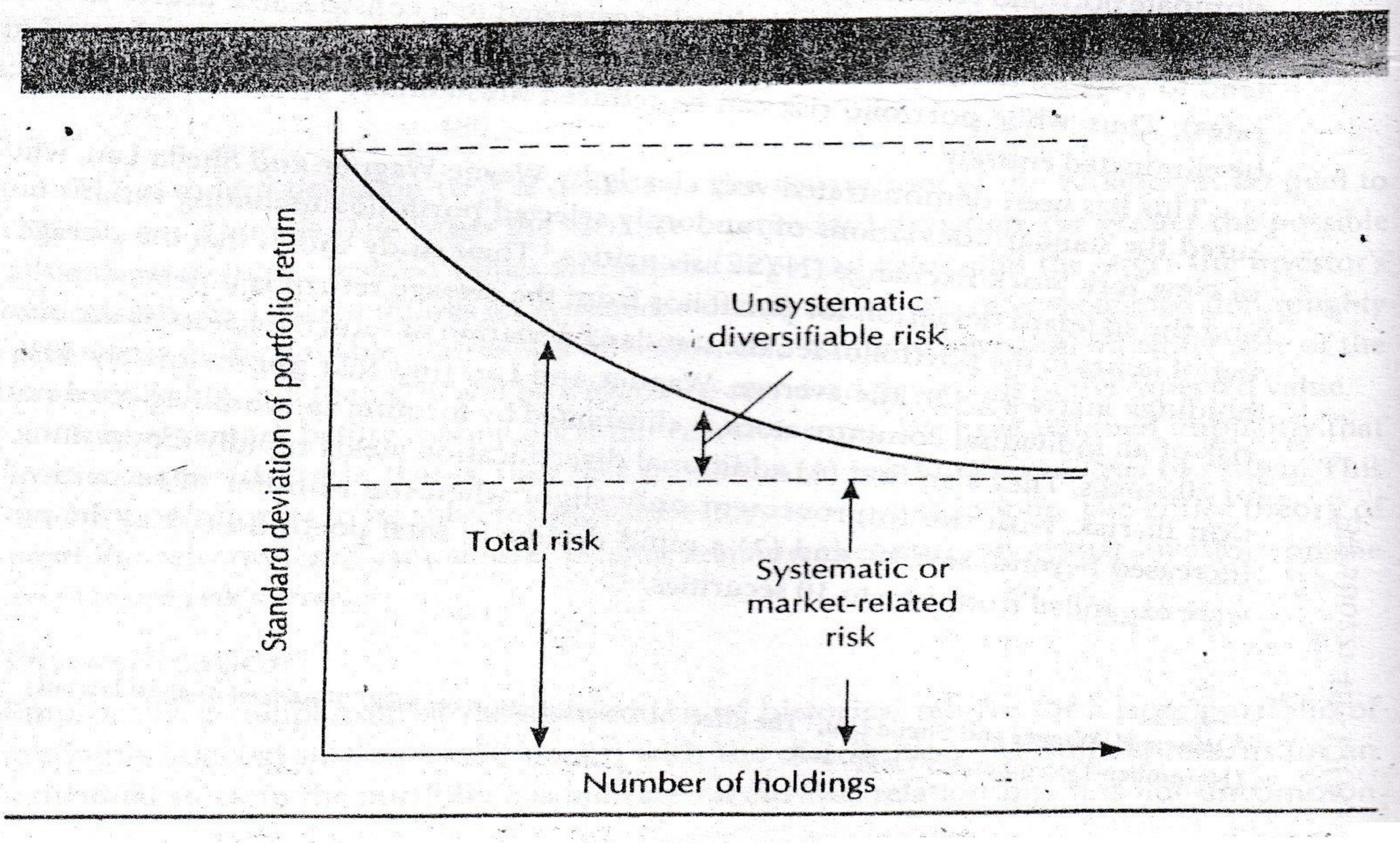
This has been demonstrated very clearly by Wayne Wagner and Sheila Lau, who measured the standard deviations of randomly selected portfolios including various numbers of New York Stock Exchange (NYSE) securities.[4] Their study shows that the average return and the standard deviation for portfolios from the average return is unrelated to the number of issues in the portfolio. Yet, the standard deviation of return declines as the number of holdings increases. On the average, Wagner and Lau find that approximately 40% of the risk of an individual common stock is eliminated by forming randomly selected portfolios of 20 stocks. They also find (1) additional diversification yields rapidly diminishing reduction in risk, with the improvement only slight when the number of securities held is increased beyond, say, 10, and (2) a rapid decline in total portfolio risk as the portfolios were expanded from one to 10 securities.

---

[4] Wayne H. Wagner and Sheila Lau, "The Effect of Diversification on Risk," *Financial Analysts Journal* (November–December 1971), pp. 2–7.

Another key finding of the Wagner-Lau study was that the return on a diversified portfolio follows the market very closely. The degree of association is measured by the correlation coefficient of each portfolio with an unweighted index of NYSE stocks. Two securities with perfectly correlated patterns will have a correlation coefficient of 1.0. Conversely, if the return patterns are perfectly negatively correlated, the correlation coefficient will equal −1.0. Two securities with uncorrelated (that is, statistically unrelated) returns will have a correlation coefficient of zero. The 20-security portfolio in the Wagner-Lau study had a correlation of 0.89 with the market. The implication is that the risk remaining in the 20-stock portfolio is predominantly a reflection of uncertainty about the performance of the stock market in general.

These results of the Wagner-Lau study show that, while some risks can be eliminated through diversification, others cannot. Thus, we are led to distinguish between a security's **unsystematic risk**, which can be washed away by mixing the security with other securities in a diversified portfolio, and its *systematic risk*, which cannot be eliminated by diversification. As the Wagner-Lau study shows, total portfolio risk declines as the number of holdings increases. Increasing diversification gradually tends to eliminate the unsystematic risk, leaving only systematic, that is, market-related risk. The remaining variability results from the fact that the return on nearly every security depends to some degree on the overall performance of the market. This is illustrated in Figure 2. Consequently, the return on a well-diversified portfolio is highly correlated with the market, and its variability or uncertainty is basically the uncertainty of the market as a whole. Investors are exposed to market uncertainty no matter how many stocks they hold.



The Risk of Individual Securities

From the empirical evidence we have cited, we can conclude that the systematic risk of an individual security is that portion of its total risk (which, again, is the same as standard deviation of return) that cannot be eliminated by combining it with other securities in a well-diversified portfolio. We now need a way of quantifying the systematic risk of a security and relating the systematic risk of a portfolio to that of its component securities. This can be accomplished by dividing a security's return into two parts: one perfectly correlated with and proportionate to the market return, and a second independent from (uncorrelated with) the market. The first component of return is usually referred to as *systematic*, the second as *unsystematic*. Thus, we have

$$\text{Security return} = \text{Systematic return} + \text{Unsystematic return} \tag{4}$$

Because the systematic return is proportional to the market return, it can be expressed as the symbol beta (β) times the market return, R_m. The proportionality factor of beta is a market sensitivity index, indicating how sensitive the security return is to changes in the market level. (How to estimate beta for a security or portfolio is discussed later.) The unsystematic return, which is independent of market returns, is usually represented by the symbol *epsilon prime* (ϵ'). Thus, the security return, R, may be expressed

$$R = \beta R_m + \epsilon' \tag{5}$$

For example, if a security has a β factor of 2.0, then a 10% market return will generate a 20% systematic return for the stock. The security return for the period would be the 20% plus the unsystematic component.

The unsystematic component in equation (5) depends on factors that are unique or specific to the company that issued the security, such as labor difficulties, unexpectedly high or low sales, managerial issues, and so on. Thus, the unsystematic component is not linked to the overall market or economic system. It is important to note that a security's return variability that is due to this unsystematic component is diversifiable precisely because that variability represents events or situations that are specific (or unique) to the firm that issued the security. If a portfolio contains many securities issued by many companies, then events that are specific to any firm and affect that firm's returns can readily be offset by contrary or opposite developments that are unique to other firms and that influence their returns in different ways.

The security returns model given by equation (5) is usually written in such a way that the average value of the residual term, ϵ', is zero. This is accomplished by adding a factor, **alpha** (α), to the model to represent the average value of the unsystematic returns over time. That is, we set $\epsilon' = \alpha + \epsilon$ so that

$$R = \alpha + \beta R_m + \epsilon \tag{6}$$

where the average ϵ over time should tend to zero. The reason the average value of the term is zero is that the term represents the eventual impact of unexpected events that are particular to the firm. If those events are random and unpredictable, their overall impact should be zero.

The model for security returns given by equation (6) is usually referred to as the **market model.**[5] Graphically, the model can be depicted as a line fitted to a plot of security returns against rates of return on the market index. This kind of line fitting is shown for a hypothetical security in Figure 3.

The beta factor can be thought of as the slope of the line. It gives the expected increase in security return for a 1% increase in market return. In Figure 3, the security has a beta of 1.0. Thus, a 10% market return will result, on the average, in a 10% security return.

The alpha factor is represented by the intercept of the line on the vertical security return axis. It is equal to the average value over time of the unsystematic returns (ϵ') on the stock. For most stocks, the alpha factor tends to be small and unstable.

Using this definition of security return given by the market model, the specification of systematic and unsystematic risk is simply the standard deviations of the two return components.

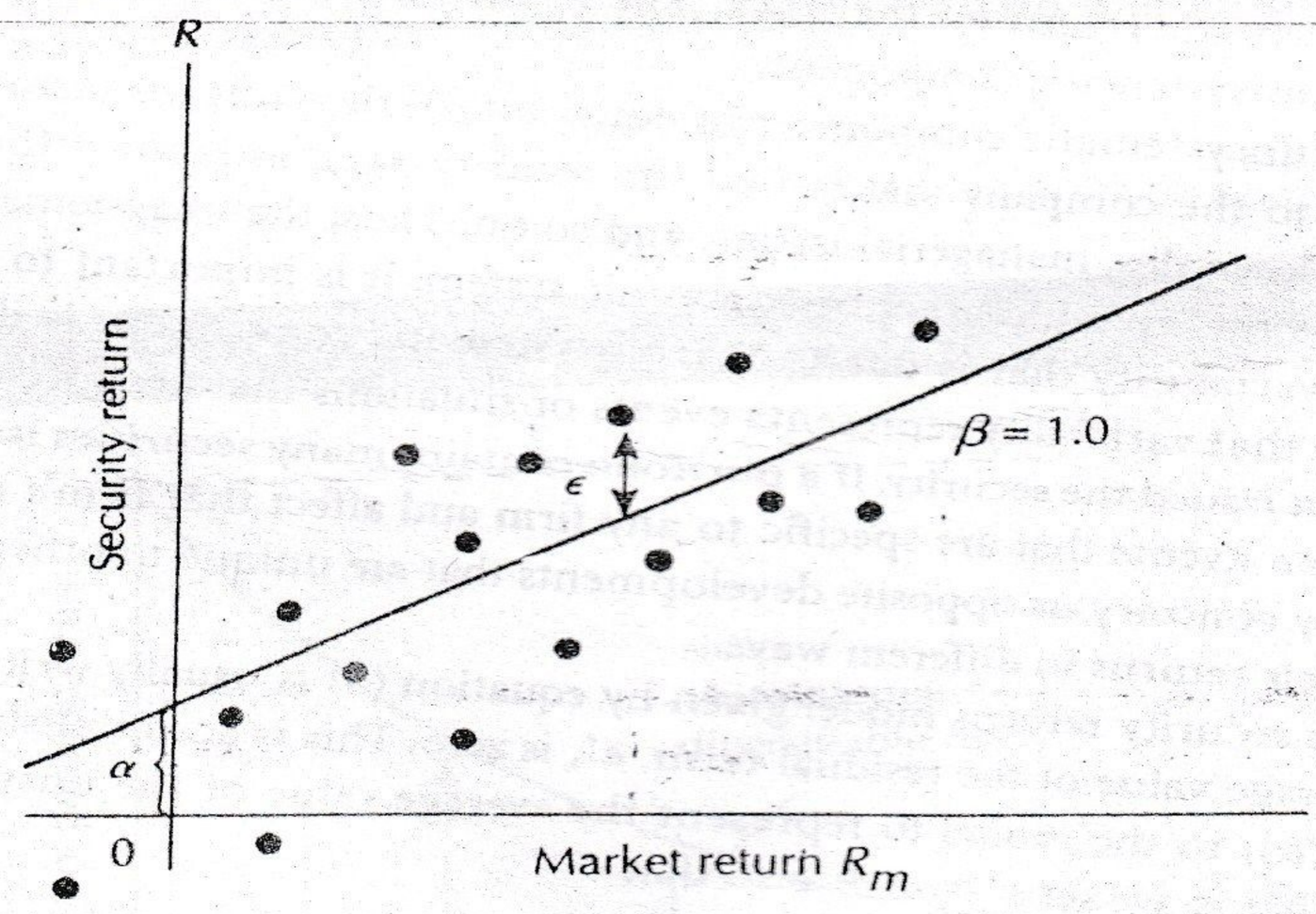
The systematic risk of a security is equal to β times the standard deviation of the market return:

$$\text{Systematic risk} = \beta\sigma_m \tag{7}$$

The unsystematic risk equals the standard deviation of the residual return factor ϵ, or

$$\text{Unsystematic risk} = \sigma_\epsilon \tag{8}$$

Figure 3 The Market Model for Security Returns



Beta (β), the market sensitivity index, is the slope of the line.
Alpha (α), the average of the residual return, is the intercept of the line on the security axis.
Epsilon (ϵ), the residual returns, are the perpendicular distances of the points from the line.

[5] It is also referred to as the *single-index market model* and the *characteristic line.*

Given measures of individual-security systematic risk, we can now compute the systematic risk of the portfolio. It is equal to the beta factor for the portfolio, β_p, times the risk of the market index, σ_m:

$$\text{Portfolio systematic risk} = \beta_p \sigma_m \tag{9}$$

The portfolio beta factor in turn can be shown to be simply an average of the individual security betas, weighted by the proportion of each security in the portfolio, or

$$\beta_p = X_1\beta_1 + X_2\beta_2 + \cdots + X_n\beta_n$$

or more concisely as

$$\beta_p = \sum_{i=1}^{n} X_i\beta_i \tag{10}$$

where

X_i = the proportion of portfolio market value represented by security i
n = the number of securities

Thus, the systematic risk of a portfolio is simply the market value-weighted average of the systematic risk of the individual securities. It follows that the β for a portfolio consisting of all stocks is 1. If a stock's β exceeds 1, it is above the average; if its beta is below 1, it is below the average. If the portfolio is composed of an equal dollar investment in each stock, the β_p is simply an unweighted average of the component security betas.

The unsystematic risk of the portfolio is also a function of the unsystematic security risks, but the form is more complex. The important point is that, with increasing diversification, this risk approaches zero.

To summarize these results, first, roughly 40% to 50% of total security risk can be eliminated by diversification. Second, the remaining systematic risk is equal to the security β times market risk. Third, portfolio systematic risk is a weighted average of security systematic risks.

The implications of these results are substantial. First, we would expect realized rates of return over substantial periods of time to be related to the systematic as opposed to the total risk of securities. As the unsystematic risk is relatively easily eliminated, we should not expect the market to offer investors a risk premium for bearing such risk. Second, because security systematic risk is equal to the security beta times σ_m (which is common to all securities), beta is useful as a *relative* risk measure. The β gives the systematic risk of a security or portfolio) relative to the risk of the market index. Thus, it is often convenient to speak of systematic risk in relative terms, that is, in terms of beta rather than beta times σ_m.

Estimating Beta

The beta of a security or portfolio can be estimated using statistical analysis. More specifically, we use regression analysis on historical data to estimate the market model given by equation (6). The estimated slope for the market model is the estimate of beta. A series of returns is computed according to equation (1) over some time interval for some broad market index (such as the S&P 500 stock market index) and for the stock (or portfolio). For example, monthly returns can be calculated for the past five years; thus, there would be 60 return observations for both the market index and the stock or portfolio. Or weekly returns can be calculated for the past year.

There is nothing in portfolio theory that indicates whether weekly, monthly, or even daily returns should be used. Nor does theory indicate any specific number of observations, except that statistical methodology entails that more observations will give a more reliable measure of beta.[6] Our purpose here is not to provide an explanation of the mechanics of calculating beta but to point out the practical problems in obtaining beta. (There are many statistical issues also, but we do not focus on these.)

There will be a difference in the calculated beta depending on: (1) the length of time over which a return is calculated (e.g., daily, weekly, monthly), (2) the number of observations used (e.g., three years of monthly returns or five years of monthly returns), (3) the specific time period used (January 1, 1996, to December 31, 2000, for example, or January 1, 1995, to December 31, 1999); and (4) the market index selected (for instance, the S&P 500 stock market index or an index consisting of all stocks traded on exchanges weighted by their relative market value). Moreover, there is the question of the stability of beta over different time intervals—that is, does the beta of a stock or portfolio remain relatively unchanged over time, or does it change?

A very interesting question has to do with the economic determinants of the beta of a stock. The risk characteristics of a company should be reflected in its beta. Several empirical studies have attempted to identify these macroeconomic and microeconomic factors.

Key Points That You Should Understand Before Proceeding

1. Risk aversion means that investors want to minimize risk for any given level of expected return, or want to maximize return, for any given level of risk.
2. Risk is the likelihood that an actual return will deviate from the expected return.
3. Investors hold diversified portfolios rather than individual securities because portfolios eliminate some of the risk of most securities.
4. The relevant risk of any individual security is not total variability in returns but rather systematic variability, which is that portion of total variability that cannot be eliminated by combining it with other securities in a diversified portfolio.
5. The market model is the hypothesis that a security's return may be attributed to two forces, the returns on securities in general or the market, and events related to the firm itself.
6. The index of the sensitivity of a security's returns to movement in the market is the security's beta, which can be estimated with regression techniques from historical data.

THE CAPITAL ASSET PRICING MODEL

We have now developed two measures of risk: one is a measure of total risk (standard deviation), the other a relative index of systematic or nondiversifiable risk (beta). The beta measure would appear to be the more relevant for the pricing of securities. Returns expected by investors logically should be related to systematic as opposed to total risk. Securities with higher systematic risk should have higher expected returns.

[6] This assumes that the economic determinants that affect the beta of a stock do not change over the measurement period.

The question of interest now is the form of the relationship between risk and return. In this section we describe a relationship called the **capital asset pricing model** (CAPM), which is based on elementary logic and simple economic principles.[7] The basic postulate underlying this finance theory is that assets with the same systematic risk should have the same expected rate of return; that is, the prices of assets in the capital markets should adjust until equivalent risk assets have identical expected returns. This principle is called the "law of one price."

To see the implications of this postulate, consider an investor who holds a risky portfolio that has the same risk as the market portfolio (beta equal to 1.0).[8] What return should she expect? Logically, she should expect the same return as that of the market portfolio. Consider another investor who holds a riskless protfolio, or one with a beta equal to zero. In this case, the investor should expect to earn the rate of return on riskless assets, such as Treasury bills. In other words, the investor who takes no risk earns only the riskless rate of return.

Now, consider the case of an investor who holds a mixture of these two portfolios. Assume he invests a proportion X of his money in the risky portfolio and the rest, or $(1 - X)$, in the riskless portfolio. What systematic risk does he bear, and what return should he expect? The risk of the composite portfolio is easily computed. Recall that the beta of a portfolio is simply a weighted average of the component security betas, where the weights are the portfolio proportions. Thus, the portfolio beta, β_p, is a weighted average of the beta of the market portfolio and that of the riskless portfolio. The market beta is 1.0, and the beta of the risk-free asset is 0. Thus,

$$\begin{aligned} \beta_p &= (1 - X) \times 0 + X \times 1 \\ &= X \end{aligned} \tag{11}$$

Thus, β_p is equal to the fraction of the money invested in the risky portfolio. If 100% or less of the investor's funds are invested in the risky portfolio, the portfolio beta will be between 0 and 1. If the investor borrows at the risk-free rate and invests the proceeds in the risky portfolio, so that X is larger than 1 and $(1 - X)$ is negative, the portfolio beta will be greater than 1.

The expected return of the composite portfolio is also a weighted average of the expected returns on the two portfolios; that is,

$$\mathrm{E}(R_p) = (1 - X) \times R_f + X \times \mathrm{E}(R_m) \tag{12}$$

where $\mathrm{E}(R_p)$ and $\mathrm{E}(R_m)$ are the expected returns on the portfolio and the market index, and R_f is the risk-free rate. Now, from equation (11) we know that X is equal to β_p. Substituting into equation (12), we have

$$\mathrm{E}(R_p) = (1 - \beta_p) \times R_f + \beta_p \times \mathrm{E}(R_m) \tag{13}$$

or

$$\mathrm{E}(R_p) = R_f + \beta_p[\mathrm{E}(R_m) - R_f]$$

---

[7] CAPM theory was developed by William F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance* (September 1964), pp. 425–442.

[8] We use the term *portfolio* in a general sense, including the case where the investor holds only one security. Because portfolio return and (systematic) risk are simply weighted averages of security values, risk/return relationships that hold for securities must also be true for portfolios, and vice versa.

Equation (13) is the capital asset pricing model. This extremely important theoretical result says that the expected return on a portfolio should exceed the riskless rate of return by an amount that is proportional to the portfolio beta. That is, the relationship between expected return and risk should be linear.

The CAPM is often stated in risk premium form. Risk premiums or excess returns are obtained by subtracting the risk-free rate from the rate of return. The expected portfolio and market risk premiums—designated $E(r_p)$ and $E(r_m)$, respectively—are given by

$$E(r_p) = E(R_p) - R_f$$

and

$$E(r_m) = E(R_m) - R_f$$

Substituting these risk premiums into equation (13), we obtain

$$E(r_p) = \beta_p E(r_m) \tag{14}$$

In this form, the CAPM states that the expected risk premium for the investor's portfolio is equal to its beta value times the expected market risk premium. Or, equivalently stated, the expected risk premium should be equal to the *quantity of risk* (as measured by beta) and the *market price of risk* (as measured by the expected market risk premium).

We can illustrate the model by assuming that the short-term (risk-free) interest rate is 6% and the expected return on the market is 10%. The expected risk premium for holding the market portfolio is simply the difference between the 10% and the short-term interest rate of 6%, or 4%. Investors who hold the market portfolio expect to earn 10%, which is 4% greater than they could earn on a short-term market instrument for certain. In order to satisfy equation (13), the expected return on securities or portfolios with different levels of risk must be as follows:

Beta	Expected Return
0.0	6%
0.5	8
1.0	10
1.5	12
2.0	14

The predictions of the model are inherently sensible. For safe investments ($\beta = 0$), the model predicts that investors would expect to earn the risk-free rate of interest. For a risky investment, ($\beta > 0$), investors would expect a rate of excess return proportional to the market sensitivity (or β) of the investment. Thus, stocks with lower-than-average market sensitivities would offer expected returns less than the expected market return. Stocks with above-average values of beta would offer expected returns in excess of the market.

Underlying Assumptions

In our development of the CAPM, we have implicitly made a number of assumptions that are required if the model is to be established on a rigorous basis. These assumptions involve

investor behavior and conditions in the capital markets. The following assumptions are sufficient to allow a single derivation of the model:

1. The market is made up of risk-averse investors who measure risk in terms of standard deviation of portfolio return. This assumption provides a basis for the use of risk measures such as beta.
2. All investors have a common time horizon for investment decision making (for example, one month, one year, and so on). This assumption allows us to measure investor expectations over some common interval, thus making comparisons meaningful.
3. All investors are assumed to have the same expectations about future security returns and risks. The only reason why they choose different portfolios is differences in systematic risk and in risk preferences. Without this assumption, the analysis would become much more complicated.
4. Capital markets are perfect in the sense that all assets are completely divisible, there are no transactions costs or differential taxes, and borrowing and lending rates are equal to each other and the same for all investors. Without these conditions, there would exist frictional barriers to the equilibrium conditions on which the model is based.

Although these assumptions are sufficient to derive the model, it is not clear that all are necessary in this exact form. It may well be that several of the assumptions can be relaxed substantially without major change in the form of the model. A good deal of research has been conducted toward this end. As explained later in this chapter, the assumptions above dealing with investor behavior have been criticized precisely by proponents of behavioral finance.

Tests of the Capital Asset Pricing Model

The CAPM is indeed a simple and elegant model, but these qualities do not in and of themselves guarantee that it will be useful in explaining observed risk/return patterns. Here we briefly review the empirical literature on attempts to verify the model.

The major difficulty in testing the CAPM is that the model is stated in terms of investors' expectations and not in terms of realized returns. To test the CAPM, it is necessary to convert the theoretical CAPM given by equation (14) into a form that can be tested empirically. We will not go through this exercise here, but will simply provide a general statement of the model that is typically tested.[9] Nor will we delve into the statistical problems associated with testing the CAPM, although we will discuss later an important theoretical issue that raises serious questions about the testability of the CAPM and, therefore, the empirical findings of researchers.

The empirical analogue of equation (14) asserts that, over the period of time analyzed, (1) there is a linear relationship between the average risk premium return on the market and the average risk premium return on a stock or portfolio, and its slope is β, and (2) the linear relationship should pass through the origin. Moreover, according to the CAPM, beta is a complete relative measure of a stock's risk. Consequently, various alternative risk measures that might be proposed, the most common being the standard deviation of return, should not be significant contributors to the explanation of a stock's return. Recall that the standard deviation measures a stock's total risk and includes both systematic and unsystematic components.

---

[9] The interested reader can find the procedure for developing the empirical model tested in Franco Modigliani and Gerald A. Pogue, "Introduction to Risk and Return: Concepts and Evidence: Part II," *Financial Analysts Journal* (May–June 1974), pp. 69–86.

The CAPM theoretically applies to both individual securities and portfolios. Therefore, the empirical tests can be based on either. However, there are statistical problems associated with estimating the magnitude of the risk/return trade-off using individual securities. By grouping securities into portfolios, we can eliminate most of the statistical problems encountered for individual stocks and thereby get a much clearer view of the relationship between return and systematic risk.

The major results of the empirical tests conducted are summarized below:[10]

1. The evidence shows a significant positive relationship between realized returns and systematic risk as measured by beta. The average market risk premium estimated is usually less than that predicted by the CAPM, however.
2. The relationship between risk and return appears to be linear. The studies give no evidence of significant curvature in the risk/return relationship.
3. Tests that attempt to discriminate between the effects of systematic and unsystematic risk do not yield definitive results. Both kinds of risk appear to be positively related to security returns, but there is substantial support for the proposition that the relationship between return and unsystematic risk is at least partly spurious—that is, partly a reflection of statistical problems rather than the true nature of capital markets.

Obviously, we cannot claim that the CAPM is absolutely right. On the other hand, the early empirical tests do support the view that beta is a useful relative risk measure and that high-beta stocks tend to be priced so as to yield correspondingly high rates of return.

In 1977, however, Richard Roll wrote a paper criticizing previously published tests of the CAPM.[11] Roll argued that while the CAPM is testable in principle, no correct test of the theory had yet been presented. He also argued that there was practically no possibility that a correct test would ever be accomplished in the future.

The reasoning behind Roll's assertions is based on his observation that there is only one potentially testable hypothesis associated with the CAPM, namely, that the true market portfolio is mean-variance efficient. (This means that the market portfolio must have minimum risk for its level of return.) Furthermore, because the true market portfolio must contain all worldwide assets, the value of most of which cannot be observed (for example, human capital), the hypothesis is in all probability untestable.

Since 1977, there have been a number of studies that purport either to support or reject the CAPM. These tests have attempted to examine implications of the CAPM other than the linearity of the risk/return relation as the basis of their methodology. Unfortunately, none provides a definitive test, and most are subject to substantial criticism, suffering from the same problem of identifying the "true" market portfolio.

---

[10] Some of the earlier studies are Nancy Jacob, "The Measurement of Systematic Risk for Securities and Portfolios: Some Empirical Results," *Journal of Financial and Quantitative Analysis* (March 1971), pp. 815–834; Merton H. Miller and Myron S. Scholes, "Rates of Returns in Relation to Risk: A Reexamination of Recent Findings," and Fischer Black, Michael C. Jensen, and Myron S. Scholes, "The Capital Asset Pricing Model: Some Empirical Evidence," in Michael C. Jensen (ed.), *Studies in the Theory of Capital Markets* (New York: Praeger Books, 1972); Marshall E. Blume and Irwin Friend, "A New Look at the Capital Asset Pricing Model," *Journal of Finance* (March 1973), pp. 19–33; and Eugene F. Fama and James D. MacBeth, "Risk, Return and Equilibrium: Empirical Tests," Working Paper No. 7237, University of Chicago, Graduate School of Business, August 1972.

[11] Richard Roll, "A Critique of the Asset Pricing Theory: Part I. On the Past and Potential Testability of the Theory," *Journal of Financial Economics* (March 1977), pp. 129–176.

Key Points That You Should Learn from This Section

1. The CAPM hypothesizes that assets with the same level of systematic risk should experience the same level of returns.
2. The level of returns expected from any asset (which may be an individual security or a portfolio of securities) is a linear function of the risk-free rate, the asset's beta, and the returns expected on the market portfolio of risky assets.
3. Some reservations about the CAPM are inevitable because it makes many assumptions about investors' behavior and the structure of the market where assets are traded.
4. A major criticism of the CAPM is that it is basically untestable because the true or relevant market portfolio is an unobservable or unattainable portfolio diversified across all risky assets in the world.

THE MULTIFACTOR CAPITAL ASSET PRICING MODEL

The CAPM described above assumes that the only risk that an investor is concerned with is uncertainty about the future price of a security. Investors, however, usually are concerned with other risks that will affect their ability to consume goods and services in the future. Three examples would be the risks associated with future labor income, the future relative prices of consumer goods, and future investment opportunities.

Recognizing these other risks that investors face, Robert Merton has extended the CAPM to describe consumers deriving their optimal lifetime consumption when they face these "extra-market" sources of risk.[12] These extra-market sources of risk are also referred to as *factors*, so the model derived by Merton is called a *multifactor CAPM* and is given below in risk premium form:

$$E(r_p) = \beta_{pm}E(r_m) + \beta_{pF1}E(r_{F1}) + \beta_{pF2}E(r_{F2}) + \cdots + \beta_{pFk}E(r_{Fk}) \qquad (15)$$

where

K = number of factors or extra-market sources of risk
β_{pFk} = the sensitivity of the portfolio to the kth factor
$E(r_{Fk})$ = the expected return of factor k minus the risk-free rate

The total extra-market sources of risk are equal to

$$\beta_{pF1}E(r_{F1}) + \beta_{pF2}E(r_{F2}) + \cdots + \beta_{pFk}E(r_{Fk}) \qquad (16)$$

Equation (15) says that investors want to be compensated for the risk associated with each source of extra-market risk, in addition to market risk. Note that, if there are no extra-market sources of risk, equation (15) reduces to the CAPM as given by equation (14). In the case of the CAPM, investors hedge the uncertainty associated with future security prices through diversification by holding the market portfolio, which can be thought of as a

[12] Robert C. Merton, "An Intertemporal Capital Asset Pricing Model," *Econometrica* (September 1973), pp. 867–888.

mutual fund that invests in all securities based on their relative capitalizations. In the multifactor CAPM, besides investing in the market portfolio, investors will also allocate funds to something equivalent to a mutual fund that hedges a particular extra-market risk. While not all investors are concerned with the same sources of extra-market risk, those that are concerned with a specific extra-market risk will basically hedge them in the same way.

As individual securities are nothing more than portfolios consisting of only one security, equation (15) must hold for each security, i. That is,

$$E(r_i) = \beta_{im}E(r_m) + \beta_{iF1}E(r_{F1}) + \beta_{iF2}E(r_{F2}) + \cdots + \beta_{iFK}E(r_{FK}) \qquad (17)$$

From an empirical perspective, it may be difficult to identify the relevant extra-market risks. Moreover, it is difficult to distinguish the multifactor CAPM empirically from the next risk and return model described.

Key Points That You Should Understand Before Proceeding

1. The multifactor CAPM posits that extra-market factors influence expected returns on securities or portfolios.
2. This approach entails that a security's return has a beta-like sensitivity to each factor.

ARBITRAGE PRICING THEORY MODEL

An alternative model to the CAPM and the multifactor CAPM was developed by Stephen Ross in 1976.[13] This model is based purely on arbitrage arguments, and hence is called the **arbitrage pricing theory (APT) model**. It postulates that a security's expected return is influenced by a variety of factors, as opposed to just the single market index of the CAPM.

The APT model assumes that there are several factors that determine the rate of return on a security, not just one as in the case of the CAPM. To understand this, look back at equation (5), which states that the return on a security is dependent on its market sensitivity index and an unsystematic return. The APT in contrast states that the return on a security is linearly related to H factors. The APT does not specify what these factors are, but it is assumed that the relationship between security returns and the factors is linear.

For now, to illustrate the APT model, let's assume a simple world with a portfolio consisting of three securities in the presence of two factors. The notation is as follows:

$\tilde{R}_i$ = the random rate of return on security i ($i = 1, 2, 3$)
$E(R_i)$ = the expected return on security i ($i = 1, 2, 3$)
F_h = the hth factor that is common to the returns of all three assets ($h = 1, 2$)
β_{ih} = the sensitivity of the ith security to the hth factor
e_i = the unsystematic return for security i ($i = 1, 2, 3$)

The APT model asserts that the random rate of return on security i is given by the relationship

$$\tilde{R}_i = E(R_i) + \beta_{i1}\tilde{F}_1 + \beta_{i2}\tilde{F}_2 + \tilde{e}_i \qquad (18)$$

[13] Stephen A. Ross, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* (December 1976), pp. 343–362.

For equilibrium to exist among these three assets, an arbitrage condition must be satisfied: Using no additional funds (wealth) and without increasing risk, it should not be possible, on average, to create a portfolio to increase return. In essence, this condition states that there is no "money machine" available in the market. Ross has shown that the following risk and return relationship, in risk premium form, will result for each security i:

$$E(r_i) = \beta_{iF1}E(r_{F1}) + \beta_{iF2}E(r_{F2}) \tag{19}$$

where

r_i = the excess return of security i over the risk-free rate
β_{iFj} = the sensitivity of security i to the jth factor
r_{Fj} = the excess return of the jth systematic factor over the risk-free rate, which can be thought of as the price (or risk premium) for the ith systematic risk

Equation (19) can be generalized to the case where there are H factors as follows:

$$E(r_i) = \beta_{iF1}E(r_{F1}) + \beta_{iF2}E(r_{F2}) + \cdots + \beta_{iFH}E(r_{FH}) \tag{20}$$

Equation (20) is the APT model. It states that investors want to be compensated for all the factors that systematically affect the return of a security. The compensation is the sum of the product of the quantity of systematic risk accepted for factor i, which is measured by the beta of the security with respect to the factor, and how the financial market prices that factor's risk, which is measured by the difference between the expected return for the factor and the risk-free rate. As in the case of the two other risk and return models described earlier, an investor is not compensated for accepting unsystematic risk.

Contrast equation (20) with the CAPM. If there is only one factor, equation (20) reduces to equation (14), and the one factor would be market risk. The CAPM is, therefore, a special case of the APT model. Now, contrast equation (20) with the multifactor CAPM given by equation (15). They look similar. Both say that investors are compensated for accepting all systematic risk and no unsystematic risk. The multifactor CAPM states that one of these systematic risks is market risk, while the APT model does not.

Supporters of the APT model argue that it has several major advantages over the CAPM or the multifactor CAPM. First, it makes less restrictive assumptions about investor preferences toward risk and return. The CAPM theory assumes investors trade off between risk and return solely on the basis of the expected returns and standard deviations of prospective investments. The APT, on the other hand, simply requires that some rather unobtrusive bounds be placed on potential investor utility functions. Second, no assumptions are made about the distribution of security returns. Finally, as the APT does not rely on identifying the true market index, the theory is potentially testable.

Empirical Evidence

To date, attempts to test the APT model empirically in the stock market have been inconclusive. Indeed, because of an inability to find a set of factors that consistently explains security returns, some question whether or not the APT is testable at all. The APT gives no direction as to the choice of the factors themselves or even how many factors might be required. Thus, the APT replaces the problem of the market portfolio in the CAPM with uncertainty over the choice and measurement of the underlying factors.

One study by Nai-fu Chen, Richard Roll, and Stephen Ross suggests four plausible economic factors:[14]

1. unanticipated changes in industrial production
2. unanticipated changes in the spread between the yield on low-grade and high-grade bonds
3. unanticipated changes in interest rates and the shape of the yield curve
4. unanticipated changes in inflation

It is interesting to note that one study expands that list to five factors, some similar to the previous group and some new.[15] The factors whose unexpected changes are said to have systematic impact on stock returns are these:

1. the business cycle, as measured by the Index of Industrial Production
2. interest rates, as given by the yields on long-term government bonds
3. investor confidence, as proxied by the spread between yields on high-grade and yields on low-grade bonds that are similar with regard to call provisions, maturity, and other features
4. short-term inflation, as measured by month-to-month changes in the Consumer Price Index
5. inflationary expectations, as represented in the changes in the short-term risk-free interest rate

As this study points out, some stocks are responsive to some factors, while other stocks are closely linked to different factors. Portfolio building involves identifying the pertinent factors and finding the price/risk reward for them. In its consideration of differing sensitivities to various systematic variables, the APT departs significantly from the CAPM.

Key Points That You Should Understand Before Proceeding

1. The APT model postulates that a security's return is a function of several factors and the security's sensitivity to changes in each of them.
2. If a security's market price were to deviate from the level justified by these factors and the security's price sensitivity to them, investors would engage in arbitrage and drive the market price to an appropriate level.
3. An appealing feature of the APT model is that it makes few assumptions about investors and the structure of the market.

ATTACKS ON THE THEORY

In this chapter, we have presented standard portfolio theory about how investors make investment decisions in order to construct an efficient portfolio. This theory is a **normative** theory—a theory that describes a norm of behavior that investors should pursue in constructing a portfolio. It does not mean that it is a theory that is actually followed. That is, standard portfolio theories should make decisions based on expected return and variance of return, but there is no assurance that they do. Also, the use of variance of returns depends

---

[14] Nai-fu Chen, Richard Roll, and Stephen A. Ross, "Economic Forces and the Stock Market: Testing the APT and Alternative Asset Pricing Theories," *Journal of Business* (July 1980), pp. 383–403.

[15] Roll and Ross Asset Management Corporation, *APT: Balancing Risk and Return* (undated), p. 14.

on whether the return distribution is symmetric and further, whether the return distribution is normally distributed. In this section, we review two attacks on standard portfolio theory. The first is an attack on the nature of the return distribution based on what has been empirically observed for financial assets. The second is the attack by the behavioral finance camp as to how investors make decisions.

Asset Return Distribution and Risk Measures

As explained in the chapter, the selection of a risk measure when investing in a financial asset is linked to the probability distribution for the return. If the distribution is symmetric, then the standard deviation can be and is used as a measure of risk.

Despite its use in finance, the preponderance of empirical evidence does not support the assumption that asset returns follow a normal distribution.[16] The evidence suggests that not only may observed probability distributions not be symmetric (i.e, they may be skewed) but that they unquestionably exhibit fat tails or heavy tails. The "tails" of the distribution are where the extreme values occur. If a probability distribution exhibits fat tails, this means that extreme values are more likely than would be predicted by the normal distribution. This means that between periods when the market exhibits relatively modest changes in returns, there will be periods when there are changes that are much higher (i.e., crashes and booms) than the normal distribution predicts.

To illustrate this, look at Figure 4. This figure shows the daily returns on a popular stock market index, the Standard & Poor's 500 (S&P 500), from 1928 to April 2006. Shown in the figure is the normal distribution. As noted earlier in this chapter, if stock returns do follow a normal distribution, returns that are three standard deviations from the mean (the mean is zero) are highly unlikely to occur. As can be seen from the figure, a considerable number of outliers (i.e., extreme values) have occurred. Four noteworthy events are the market crashes of October 19, 1987; October 28, 1929; and October 29, 1929; and the market boom on March 15, 1933. The number of standard deviations from the mean is reported. The likelihood of an event with the number of standard deviations shown for these four events if daily returns follow a normal distribution is zero. This is just one form of empirical evidence that asset returns do not follow a normal distribution.

The initial assault on the assumption that return distributions are not normally distributed was in the 1960s by Benoit Mandelbrot.[17] Empirically investigating various time series on commodity returns and interest rates, he not only rejected the assumption that returns (and prices) followed a normal distribution, but he conjectured that they are better described by a non-normal "stable" distribution. A discussion of a stable distribution, more specifically what Mandelbrot labeled a "stable Paretian distribution," is beyond the scope of this book.[18] His early investigations on asset returns were carried out further by Eugene Fama,[19] among others, and led to a consolidation of the hypothesis that asset returns can be better described as a stable Paretian distribution.

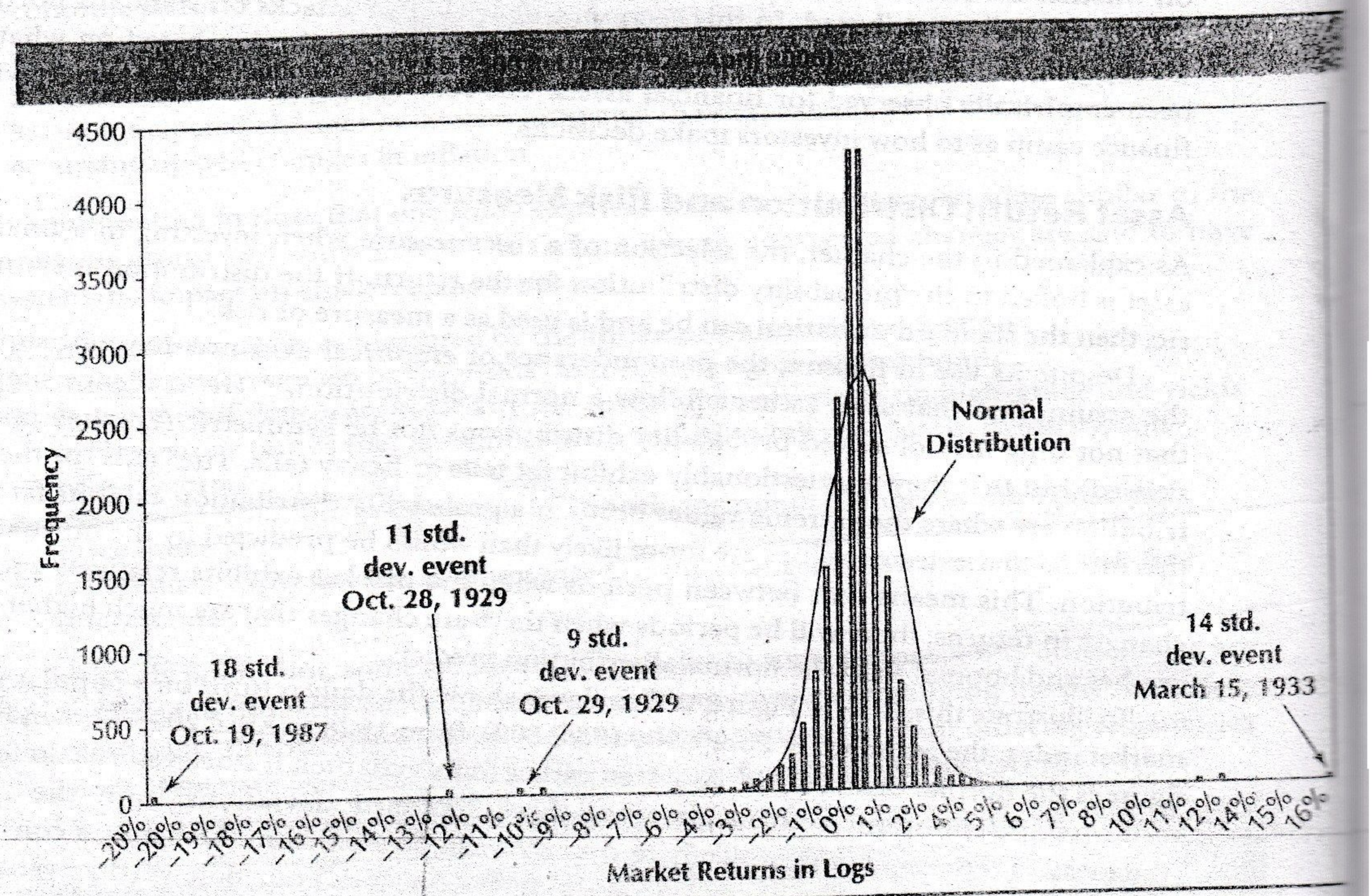
---

[16] See Chapter 11 in Svetlozar T. Rachev, Christian Menn, and Frank J. Fabozzi, *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing* (John Wiley & Sons, 2005).

[17] Benoit B. Mandelbrot, "The Variation of Certain Speculative Prices," *Journal of Business*, 36 (1963), pp. 394–419.

[18] See Chapter 7 in Rachev, Menn, and Fabozzi, *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*.

[19] Eugene F. Fama, "Mandelbrot and the Stable Paretian Hypothesis," *Journal of Business*, 36 (1963), pp. 420–429.



Source: *Exhibit 4 in Jeremy Grantham, "Risk Management in Investing (Part Two): Risk and the Passage of Time," Letters to the Investment Committee VII (April 2006). © 2006 GMO LLC.*

An important implication of asset returns following a stable Paretian distribution is that the standard deviation is infinite. As we have seen, the standard deviation plays a key role in portfolio theory as well as capital market theory, discussed in the next chapter. If the standard deviation is infinite, then how could the standard deviation be used as a measure of risk?

As a result, some researchers have suggested other probability distributions that have a finance variance to get around the infinite variance problem. For example, one alternative type of distribution is the Student's *t* distribution, a distribution that under certain conditions not only has a finite variance but also allows for tails with more observations than the normal distribution.[20] Another distribution that has been proposed is a finite mixture of normal distributions.[21] Despite these proposals, there are strong theoretical reasons in favor of the stable Paretian distribution.[22]

---

[20] Empirical evidence supporting this distribution was provided by Robert C. Blattberg and Nicholas J. Gonedes, "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices," *Journal of Business*, 47 (1974), pp. 244–280.

[21] Kon found that this distribution explains daily returns for stocks better than the Student's *t* distribution. See Stanley Kon, "Models of Stock Returns—A Comparison," *Journal of Finance*, 39 (1984), pp. 147–165.

[22] The reason is that a major drawback of all these alternative models is their lack of stability. This property is highly desirable for asset returns in the context of portfolio analysis and risk management. Only for stable (which includes the normal as a special case) distributed returns of independent assets does one obtain the property that the linear combination of the returns (portfolio returns) again follow a stable distribution. For the technical arguments, see Svetlozar T. Rachev and Stefan Mittnik, *Stable Paretian Models in Finance* (Chichester, UK: John Wiley & Sons, 2000).

Support for return distributions not being well characterized by a normal distribution comes from the collaborative efforts of economists and physicists who have developed mathematical models of the stock market.[23] In modeling the stock market, the interactions of investors are modeled. Although it is well recognized that such mathematical models are by their nature gross simplifications of real-world financial markets, they provide sufficient structure to analyze return distributions. The computer simulations of these models have been found to generate fat tails and other statistical characteristics that have been observed in real-world financial markets.

The issue of the return distribution also leads to competitors for how risk should be measured. There is a growing literature in finance in two areas. The first has been the development of the theory for assessing risk measures.[24] The second is alternative risk measures.[25] These two areas are beyond the scope of this chapter.

At this time, there is considerable research seeking to modify much of the finance theory described in this and the next chapter assuming that return distributions follow a distribution other than a normal distribution.

Assault by the Behavioral Finance Theory Camp

One of the earliest attacks on the foundations of economic principles that financial economists draw upon in formulating financial theories was in the late 1970s by Daniel Kahneman and Amos Tversky.[26] Based on numerous experiments, these two psychologists demonstrated that the actions of decision makers are inconsistent with the assumptions made by economists. Basically, they formulated a theory, known as "prospect theory," that attacks expected utility theory.[27] While we have not presented a full-blown derivation of capital market theory in this chapter, if we had, we would have seen the importance of expected utility theory.

Other attacks on the assumptions of standard financial theory drawing from the field of psychology led to the specialized field in finance known as behavioral finance.[28] **Behavioral finance** looks at how psychology affects investor decisions and the implications

---

23 Probably the most well-known model is the Santa Fe Stock Market Model. See W. Brian Arthur, John H. Holland, Blake LeBaron, Richard Palmer, and Paul Tayler, "Asset Pricing under Endogeneous Expectations in an Artificial Stock Market," in W. Brian Arthur, Steven N. Durlauf, and David A. Lane (eds.), *The Economy as an Evolving Complex System II* (Reading, MA: Addison-Wesley, 1997), pp. 15–44.

24 See Philippe Artzner, Freddy Delbaen, Jean-Marc Heath, and David Eber, "Coherent Measures of Risk," *Mathematical Finance*, 9 (2000), pp. 203–228.

25 See Sergio Ortobelli, Svetlozar T. Rachev, Stoyan Stoyanov, Frank J. Fabozzi, and Almira Biglova, "The Proper Use of Risk Measures in Portfolio Theory," *International Journal of Theoretical and Applied Finance*, 8 (December 2005), pp. 1–27, and Svetlozar T. Rachev, Sergio Ortobelli, Stoyan Stoyanov, Frank J. Fabozzi, and Almira Biglova, "Desirable Properties of an Ideal Risk Measure in Portfolio Theory," *International Journal of Theoretical and Applied Finance*, 11 (February 2008), pp. 19–54.

26 It should be noted that the economists Adam Smith and John Maynard Keynes thought that investor psychology affected security prices.

27 Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* (March 1979), pp. 236–291. See also Daniel Kahneman and Amos Tversky, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5 (1992), pp. 297–323.

28 For an extensive discussion of behavioral finance, see the following chapters in Frank J. Fabozzi (ed.), *The Handbook of Finance: Volume II* (Hoboken, N.J.: John Wiley & Sons, 2008): Meir Statman, "What Is Behavioral Finance"; Jarrod W. Wilcox, "Behavioral Finance"; Victor Ricciardi "The Psychology of Risk: The Behavioral Finance Perspective"; and Victor Ricciardi, "Risk: Traditional Finance versus Behavioral Finance."

not only for portfolio theory that we described in this chapter but also asset pricing theory and market efficiency. The foundations of behavioral finance draw from the research by Kahneman, Slovic, and Tversky.[29]

Shefrin notes the following three themes in the behavioral finance literature:[30]

Behavioral Finance Theme 1: Investors err in making investment decisions because they rely on rules of thumb.
Behavioral Finance Theme 2: Investors are influenced by form as well as substance in making investment decisions.
Behavioral Finance Theme 3: Prices in the financial market are affected by errors and decision frames.

The first behavioral finance theme involves the concept of **heuristics**. This term means a rule-of-thumb strategy or good guide to follow in order to shorten the time it takes to make a decision. For example, here are three rules of thumb provided on *MSN Money*'s website for increasing the likelihood of success when investing in common stock: (1) ignore guru predictions, (2) avoid cheap stocks, (3) follow the big players.[31] In planning for retirement, a rule of thumb that has been suggested for having enough to retire is to invest 10% of annual pretax income. As for what to invest in to reach that retirement goal (that is, the allocation among asset classes), a rule of thumb that has been suggested is that the percentage that an investor should allocate to bonds should be determined by subtracting from 100 the investor's age. So, for example, a 45-year-old individual should invest 55% of his or her retirement funds in bonds.

There are circumstances where heuristics can work fairly well. However, the psychology literature tells us that heuristics can lead to systematic biases in decision making, what psychologists refer to as **cognitive biases**. In the context of finance, these biases lead to errors in making investment decisions. Shefrin refers to these biases as **heuristic-driven biases.**

Now recall how investment decisions were assumed to be made in the standard finance theory that we described earlier in this chapter. The assumption is that an investor computes the mean and variance of the return of financial assets and constructs an optimal portfolio based on that statistical information.

The second behavioral finance theme involves the concept of framing. The term **framing** means the way in which a situation or choice is presented to an investor. Behavioral finance asserts that the framing of investment choices can result in significantly different assessments by an investor as to the risk and return of each choice and therefore the ultimate decision made.[32] Shefrin and Statman provide an example of faulty framing coupled with a cognitive bias.[33] Individual investors often fail to treat the value of their stock portfolio at market value. Instead, investors have a "mental account" where they continue to market the

---

[29] Daniel Kahneman, Paul Slovic, and Amos Tversky, *Judgment under Uncertainty: Heuristics and Biases* (New York: Cambridge University Press, 1982).
[30] Hersh Shefrin, *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing* (New York: Oxford Universtity Press, 2002), pp. 4–5.
[31] Harry Domash, "10 Rules for Picking Stock Winners," *MSN Money*, September 27, 1997.
[32] See Amos Tversky and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 211 (1961), pp. 453–458, and Amos Tversky and Daniel Kahneman, "Rational Choice and the Framing of Decisions," *Journal of Business*, 59 (October 1986), Part 2, pp. 251–278.
[33] Hersh Shefrin and Meir Statman, "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence," *Journal of Finance*, 40 (July 1985), pp. 777–790.

value of each stock in their portfolio at the purchase price despite the change in the market value. The reason why they are reluctant to acknowledge any losses on stocks that they own is that it keeps alive the hope that those stocks that have realized losses will turn around so as to realize a gain. When they finally sell their stocks, they close the mental account and only at that time acknowledge a loss that had occurred on paper. Hence, investment decisions are affected by this mental accounting treatment rather than being based on the true economic impact that an investment decision would have on the investor. Behavioral finance assumes that there is "frame dependence" when it comes to making investment decisions and hence the second theme of behavioral finance carries that name. In contrast, standard finance theory assumes "frame independence." This means that investors "view all decisions through the transparent, objective lens of risk and return."[34]

Finally, the third theme of behavioral finance involves how errors caused by heuristics and framing dependence affect the pricing of assets. According to behavioral finance, asset prices will not reflect their fundamental value because of the way investors make decisions. That is, markets will be price inefficient. Hence, Shefrin labels the third behavioral finance theme "inefficient markets."

That leaves us with two theories. Who is right, supporters of standard finance theory or supporters of behavioral finance theory? In fairness, we have not provided the responses of the supporters of standard finance theory to the criticisms of those who support behavioral finance. Nor have we presented the attacks on behavioral finance. Fortunately, David Hirshleifer provided that analysis. He describes the common objections to both approaches. He refers to the standard finance theory as the "fully rational approach" and behavioral finance as the "psychological approach."[35] A criticism of both approaches is that they can go "theory fishing" to find theories in market data to support their position. Objections to the fully rational approach are (1) that the calculations needed to implement this approach are extremely difficult to do, and (2) that the empirical evidence in the finance literature does not support rational behavior by investors. Objections to the psychological approach according to Hirshleifer are (1) that "alleged psychology biases are arbitrary," and (2) that the experiments performed by research that find alleged psychological biases are arbitrary.

SUMMARY

This chapter explains the principles of portfolio theory, a theory that deals with the construction of optimal portfolios by rational risk-averse investors, and its implications for the returns on different assets and, hence, security prices. One way to evaluate the risk of a portfolio is by estimating the extent to which future portfolio values are likely to deviate from expected portfolio return. This is measured by the variance of the portfolio's return and is called the total portfolio risk. Total portfolio risk can be broken down into two types of risk: systematic risk and unsystematic risk.

Systematic risk, also called *market risk*, is the risk that affects all securities. The beta of any security or portfolio is the relative systematic risk of the asset and is measured statistically

---

34 Shefrin, *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing*, p. 4.

35 See Table 1 in David Hirshleifer, "Investor Psychology and Asset Pricing," *Journal of Finance*, 56 (September 2001), pp. 1533–1597.

(using historical return data) by the slope of the regression between the asset's and the market's returns. The regression line estimated is called the market model. Unsystematic risk is the risk that is unique to a company, and it can be eliminated by diversifying the portfolio. Thus, systematic risk and unsystematic risk are referred to as nondiversifiable and diversifiable risk, respectively.

The CAPM is an economic theory that attempts to provide a relationship between risk and return, or, equivalently, it is a model for the pricing of risky securities. The CAPM asserts that the only relevant risk that investors will require compensation for assuming is systematic risk because that risk cannot be eliminated by diversification. Basically, CAPM says that the expected return of a security or a portfolio is equal to the rate on a risk-free security plus a risk premium. The risk premium in the CAPM is proportional to the security or the portfolio beta. More specifically, the risk premium is the product of the quantity of risk and the market price of risk, measured by beta, and the difference between the expected market return and risk-free rate, respectively.

The CAPM assumes that investors are concerned with only one risk: the risk having to do with the future price of a security. However, there are other risks, such as the capacity of investors to consume goods and services in the future. The multifactor CAPM assumes investors face such extra-market sources of risk called factors. The expected return in the multifactor CAPM is the market risk, as in the case of the basic CAPM, plus a package of risk premiums. Each risk premium is the product of the beta of the security or portfolio with respect to the particular factor and the difference between the expected return for the factor and the risk-free rate.

The APT model is developed purely on arbitrage arguments. It postulates that the expected return on a security or a portfolio is influenced by several factors. Proponents of the APT model cite its use of less restrictive assumptions as a feature that makes it more appealing than the CAPM or the multifactor CAPM. Moreover, testing the APT model does not require identification of the true market portfolio. It does, however, require empirical determination of the factors because they are not specified by the theory. Consequently, the APT model replaces the problem of identifying the market portfolio in the CAPM with the problem of choosing and measuring the underlying factors.

All things considered, the CAPM, the multifactor CAPM, and the APT model all provide interesting conceptual insights into the issues of the pricing of risk and portfolio selection in securities markets. None can be said to dominate the others in terms of theoretical content or the simplicity of empirical testing. Only the future will decide which has the best claim to the ultimate truth. Indeed, on this question, all three pricing theories will probably make valuable contributions to the development of the next generation of equilibrium models.

Challenges to the standard theory of portfolio selection in terms of selecting portfolios based on expected value and variance of returns have been twofold. The first is a challenge to the notion that the return distribution for financial assets is symmetric and follows a normal distribution. If that is not the case, the standard deviation may be a poor measure of risk. Alternative risk measures have been proposed. The second attack is from a branch of finance known as behavioral finance. This branch of finance, which draws from the field of psychology, asserts that investor behavior is far different from that postulated in standard finance theory.

KEY TERMS

Alpha
Arbitrage pricing theory (APT) model
Behavioral finance
Beta
Capital asset pricing model (CAPM)
Cognitive biases
Correlation
Diversifiable risk
Efficient (optimal) portfolio
Expected return
"Extra-market" source of risk
Framing
Fat tails
- Heavy tails
- Heuristic-driven biases
- Heuristics
- Market model
- Normative theory
- Probability distribution
- Standard deviation
- Symmetrical distribution
- Systematic risk
- Total risk
- Unsystematic risk
- Variance

QUESTIONS

A friend has asked you to help him figure out a statement he received from his broker. It seems that, at the start of last year, your friend paid $900 for a bond, and sold it at the end of the year for $890. During the year, he received a single coupon payment of $110. The statement claims that his return (not including commissions and taxes) is 11.11% for the year. Is this claim correct?

Suppose the probability distribution for the one-period return of some asset is as follows:

Return	Probability
0.20	0.10
0.15	0.20
0.10	0.30
0.03	0.25
−0.06	0.15

a. What is this asset's expected one-period return?
b. What is this asset's variance and standard deviation for the one-period return?

"A portfolio's expected return and variance of return are simply the weighted average of the expected return and variance of the individual assets." Do you agree with this statement?

4. In the January 25, 1991, issue of *The Value Line Investment Survey*, you note the following:

Company	Beta (β)
IBM	0.95
Bally Manufacturing	1.40
Cigna Corporation	1.00
British Telecom	0.60

a. How do you interpret these betas?
b. Is it reasonable to assume that the expected return on British Telecom is less than that on IBM shares?
c. "Given that Cigna Corporation has a β of 1.00, one can mimic the performance of the stock market as a whole by buying only these shares." Do you agree with this statement?

5. Assume the following:

Expected market return = 15%
Risk-free rate = 5.7%

If a security's beta is 1.3, what is its expected return according to the CAPM?

6. Professor Harry Markowitz, corecipient of the 1990 Nobel Prize in Economics, wrote the following:

> A portfolio with sixty different railway securities, for example, would not be as well

diversified as the same size portfolio with some railroad, some public utility, mining, various sorts of manufacturing, etc.

Why is this true?

7. Following is an excerpt from an article, "Risk and Reward," in *The Economist* of October 20, 1990:

> Next question: is the CAPM supported by the facts? That is controversial, to put it mildly. It is a tribute to Mr. Sharpe [cowinner of the 1990 Nobel Prize in Economics] that his work, which dates from the early 1960s, is still argued over so heatedly. Attention has lately turned away from beta to more complicated ways of carving up risk. But the significance of CAPM for financial economics would be hard to exaggerate.

a. Summarize Roll's argument on the problems inherent in empirically verifying the CAPM.
b. What are some of the other "more complicated ways of carving up risk"?

8. **a.** What are the difficulties in practice of applyi the arbitrage pricing theory model?
b. Does Roll's criticism also apply to this pricin model?
c. "In the CAPM investors should be compensa for accepting systematic risk: for the APT mode investors are rewarded for accepting both syste atic risk and unsystematic risk." Do you agree w this statement?

9. **a.** What does it mean that a return distribution a fat tail?
b. What is the implication if a return distributio is assumed to be normally distributed but is in f a fat-tailed distribution?

10. How does the behavioral finance approach diffe from the standard finance theory approach?