CHAPTER 1

Bond Fundamentals

Risk management starts with the pricing of assets. The simplest assets to study are regular, fixed-coupon bonds. Because their cash flows are predetermined, we can translate their stream of cash flows into a present value by discounting at a fixed interest rate. Thus the valuation of bonds involves understanding compounded interest, discounting, as well as the relationship between present values and interest rates.

Risk management goes one step further than pricing, however. It examines potential changes in the price of assets as the interest rate changes. In this chapter, we assume that there is a single interest rate, or yield, that is used to price the bond. This will be our fundamental risk factor. This chapter describes the relationship between bond prices and yields and presents indispensable tools for the management of fixed-income portfolios.

This chapter starts our coverage of quantitative analysis by discussing bond fundamentals. Section 1.1 reviews the concepts of discounting, present values, and future values. Section 1.2 then plunges into the price-yield relationship. It shows how the Taylor expansion rule can be used to relate movements in bond prices to those in yields. This Taylor expansion rule, however, covers much more than bonds. It is a building block of risk measurement methods based on local valuation, as we shall see later. Section 1.3 then presents an economic interpretation of duration and convexity.

The reader should be forewarned that this chapter, like many others in this handbook, is rather compact. This chapter provides a quick review of bond fundamentals with particular attention to risk measurement applications. By the end of this chapter, however, the reader should be able to answer advanced FRM questions on bond mathematics.

1.1 DISCOUNTING, PRESENT, AND FUTURE VALUE

An investor considers a zero-coupon bond that pays \$100 in 10 years. Assume that the investment is guaranteed by the U.S. government, and that there is no credit risk. So, this is a default-free bond, which is exposed to market risk only. Because the payment occurs at a future date, the current value of the investment is surely less than an up-front payment of \$100.

To value the payment, we need a **discounting factor**. This is also the **interest rate**, or more simply the **yield**. Define C_t as the cash flow at time t and the discounting factor as y. We define T as the number of periods until maturity, e.g., number of years, also known as **tenor**. The **present value** (PV) of the bond can be computed as

$$PV = \frac{C_T}{(1+y)^T} \tag{1.1}$$

For instance, a payment of $C_T = \$100$ in 10 years discounted at 6 percent is only worth \$55.84 now. So, all else fixed, the market value of zero-coupon bonds decreases with longer maturities. Also, keeping T fixed, the value of the bond decreases as the yield increases.

Conversely, we can compute the **future value** (FV) of the bond as

$$FV = PV \times (1+y)^T \tag{1.2}$$

For instance, an investment now worth $PV = \$100$ growing at 6 percent will have a future value of $FV = \$179.08$ in 10 years.

Here, the yield has a useful interpretation, which is that of an **internal rate of return** on the bond, or annual growth rate. It is easier to deal with rates of returns than with dollar values. Rates of return, when expressed in percentage terms and on an annual basis, are directly comparable across assets. An annualized yield is sometimes defined as the **effective annual rate (EAR)**.

It is important to note that the interest rate should be stated along with the method used for compounding. Annual compounding is very common. Other conventions exist, however. For instance, the U.S. Treasury market uses semiannual compounding. Define in this case y^S as the rate based on semiannual compounding. To maintain comparability, it is expressed in annualized form, i.e., after multiplication by 2. The number of periods, or semesters, is now $2T$. The formula for finding y^S is

$$PV = \frac{C_T}{(1+y^S/2)^{2T}} \tag{1.3}$$

For instance, a Treasury zero-coupon bond with a maturity of $T = 10$ years would have $2T = 20$ semiannual compounding periods. Comparing with (1.1), we see that

$$(1+y) = (1+y^S/2)^2 \tag{1.4}$$

Continuous compounding is often used when modeling derivatives. It is the limit of the case where the number of compounding periods per year increases to infinity. The continuously compounded interest rate y^C is derived from

$$PV = C_T \times e^{-y^C T} \tag{1.5}$$

where $e^{(\cdot)}$, sometimes noted as $\exp(\cdot)$, represents the exponential function.

Note that in all of these Equations (1.1), (1.3), and (1.5), the present value and future cash flows are identical. Because of different compounding periods, however, the yields will differ. Hence, the compounding period should always be stated.

Example: Using Different Discounting Methods

Consider a bond that pays \$100 in 10 years and has a present value of \$55.8395. This corresponds to an annually compounded rate of 6.00% using $PV = C_T/(1+y)^{10}$, or $(1+y) = (C_T/PV)^{1/10}$.

This rate can be transformed into a semiannual compounded rate, using $(1+y^S/2)^2 = (1+y)$, or $y^S/2 = (1+y)^{1/2} - 1$, or $y^S = ((1+0.06)^{(1/2)} - 1) \times 2 = 0.0591 = 5.91\%$. It can be also transformed into a continuously compounded rate, using $\exp(y^C) = (1+y)$, or $y^C = \ln(1+0.06) = 0.0583 = 5.83\%$.

Note that as we increase the frequency of the compounding, the resulting rate decreases. Intuitively, because our money works harder with more frequent compounding, a lower investment rate will achieve the same payoff at the end.

KEY CONCEPT

For fixed present value and cash flows, increasing the frequency of the compounding will decrease the associated yield.

EXAMPLE 1.1: FRM EXAM 2002—QUESTION 48

An investor buys a Treasury bill maturing in 1 month for \$987. On the maturity date the investor collects \$1,000. Calculate effective annual rate (EAR).

a. 17.0%
b. 15.8%
c. 13.0%
d. 11.6%

EXAMPLE 1.2: FRM EXAM 2002—QUESTION 51

Consider a savings account that pays an annual interest rate of 8%. Calculate the amount of time it would take to double your money. Round to the nearest year.

a. 7 years
b. 8 years
c. 9 years
d. 10 years

1.2 PRICE–YIELD RELATIONSHIP

1.2.1 Valuation

The fundamental discounting relationship from Equation (1.1) can be extended to any bond with a fixed cash-flow pattern. We can write the present value of a bond P as the discounted value of future cash flows:

$$P = \sum_{t=1}^{T} \frac{C_t}{(1+y)^t} \tag{1.6}$$

where: C_t = the cash flow (coupon or principal) in period t
t = the number of periods (e.g., half-years) to each payment
T = the number of periods to final maturity
y = the discounting factor per period (e.g., $y^S/2$)

A typical cash-flow pattern consists of a fixed coupon payment plus the repayment of the principal, or **face value** at expiration. Define c as the coupon *rate* and F as the face value. We have $C_t = cF$ prior to expiration, and at expiration, we have $C_T = cF + F$. The appendix reviews useful formulas that provide closed-form solutions for such bonds.

When the coupon rate c precisely matches the yield y, using the same compounding frequency, the present value of the bond must be equal to the face value. The bond is said to be a **par bond**. If the coupon is greater than the yield, the price must be greater than the face value, which means that this is a **premium bond**. Conversely, if the coupon is lower, or even zero for a zero-coupon bond, the price must be less than the face value, which means that this is a **discount bond**.

Equation (1.6) describes the relationship between the yield y and the value of the bond P, given its cash-flow characteristics. In other words, the value P can also be written as a nonlinear function of the yield y:

$$P = f(y) \tag{1.7}$$

Conversely, we can set P to the current market price of the bond, including any accrued interest. From this, we can compute the "implied" yield that will solve this equation.

Figure 1.1 describes the price-yield function for a 10-year bond with a 6% annual coupon. In risk management terms, this is also the relationship between the payoff on the asset and the risk factor. At a yield of 6%, the price is at par, $P = \$100$. Higher yields imply lower prices. This is an example of a **payoff function**, which links the price to the underlying risk factor.

Over a wide range of yield values, this is a highly nonlinear relationship. For instance, when the yield is zero, the value of the bond is simply the sum of cash

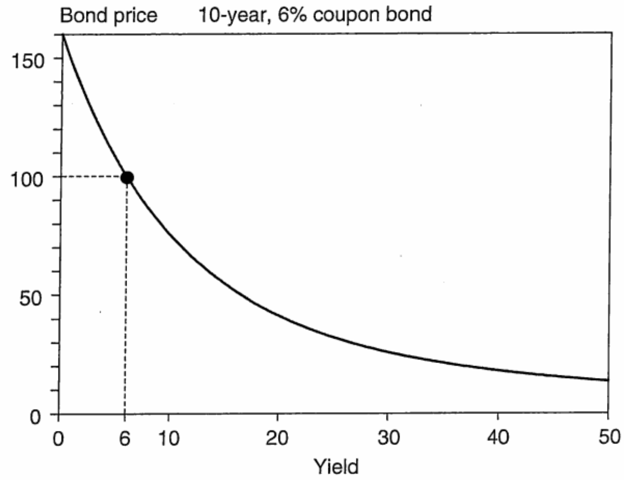

FIGURE 1.1 Price–Yield Relationship

flows, or \$160 in this case. When the yield tends to very large values, the bond price tends to zero. For small movements around the initial yield of 6%, however, the relationship is quasilinear.

There is a particularly simple relationship for **consols**, or **perpetual bonds**, which are bonds making regular coupon payments but with no redemption date. For a consol, the maturity is infinite and the cash flows are all equal to a fixed percentage of the face value, $C_t = C = cF$. As a result, the price can be simplified from Equation (1.6) to

$$P = cF\left[\frac{1}{(1+y)} + \frac{1}{(1+y)^2} + \frac{1}{(1+y)^3} + \cdots\right] = \frac{c}{y}F \tag{1.8}$$

as shown in the appendix. In this case, the price is simply proportional to the inverse of the yield. Higher yields lead to lower bond prices, and vice versa.

Example: Valuing a Bond

Consider a bond that pays \$100 in 10 years and a 6% annual coupon. Assume that the next coupon payment is in exactly one year. What is the market value if the yield is 6%? If it falls to 5%?

The bond cash flows are $C_1 = \$6$, $C_2 = \$6, \ldots, C_{10} = \$106$. Using Equation (1.6) and discounting at 6%, this gives the present value of cash flows of \$5.66, \$5.34, ..., \$59.19, for a total of \$100.00. The bond is selling at par. This is logical because the coupon is equal to the yield, which is also annually compounded. Alternatively, discounting at 5% leads to a price of \$107.72.

1.2.2 Taylor Expansion

Let us say that we want to see what happens to the price if the yield changes from its initial value, called y_0, to a new value, $y_1 = y_0 + \Delta y$. Risk management is all about assessing the effect of changes in risk factors such as yields on asset values. Are there shortcuts to help us with this?

We could recompute the new value of the bond as $P_1 = f(y_1)$. If the change is not too large, however, we can apply a very useful shortcut. The nonlinear relationship can be approximated by a **Taylor expansion** around its initial value[1]

$$P_1 = P_0 + f'(y_0)\Delta y + \frac{1}{2} f''(y_0)(\Delta y)^2 + \cdots \tag{1.9}$$

where $f'(\cdot) = \frac{dP}{dy}$ is the first derivative and $f''(\cdot) = \frac{d^2P}{dy^2}$ is the second derivative of the function $f(\cdot)$ valued at the starting point.[2] This expansion can be generalized to situations where the function depends on two or more variables. For bonds, the first derivative is related to the *duration* measure, and the second to *convexity*.

Equation (1.9) represents an infinite expansion with increasing powers of Δy. Only the first two terms (linear and quadratic) are ever used by finance practitioners. They provide a good approximation to changes in prices relative to other assumptions we have to make about pricing assets. If the increment is very small, even the quadratic term will be negligible.

Equation (1.9) is fundamental for risk management. It is used, sometimes in different guises, across a variety of financial markets. We will see later that this Taylor expansion is also used to approximate the movement in the value of a derivatives contract, such as an option on a stock. In this case, Equation (1.9) is

$$\Delta P = f'(S)\Delta S + \frac{1}{2} f''(S)(\Delta S)^2 + \cdots \tag{1.10}$$

where S is now the price of the underlying asset, such as the stock. Here, the first derivative $f'(S)$ is called *delta*, and the second $f''(S)$, *gamma*.

The Taylor expansion allows easy aggregation across financial instruments. If we have x_i units (numbers) of bond i and a total of N different bonds in the portfolio, the portfolio derivatives are given by

$$f'(y) = \sum_{i=1}^{N} x_i f_i'(y) \tag{1.11}$$

[1] This is named after the English mathematician Brook Taylor (1685–1731), who published this result in 1715. The full recognition of the importance of this result only came in 1755 when Euler applied it to differential calculus.

[2] This first assumes that the function can be written in polynomial form as $P(y + \Delta y) = a_0 + a_1\Delta y + a_2(\Delta y)^2 + \cdots$, with unknown coefficients a_0, a_1, a_2. To solve for the first, we set $\Delta y = 0$. This gives $a_0 = P_0$. Next, we take the derivative of both sides and set $\Delta y = 0$. This gives $a_1 = f'(y_0)$. The next step gives $2a_2 = f''(y_0)$. Here, the term "derivatives" takes the usual mathematical interpretation, and has nothing to do with *derivatives products* such as options.

1.3 BOND PRICE DERIVATIVES

For fixed-income instruments, the derivatives are so important that they have been given a special name.[3] The negative of the first derivative is the **dollar duration** (DD):

$$f'(y_0) = \frac{dP}{dy} = -D^* \times P_0 \tag{1.12}$$

where D^* is called the **modified duration.** Thus, dollar duration is

$$\mathrm{DD} = D^* \times P_0 \tag{1.13}$$

where the price P_0 represent the *market* price, including any accrued interest. Sometimes, risk is measured as the **dollar value of a basis point** (DVBP),

$$\mathrm{DVBP} = \mathrm{DD} \times \Delta y = [D^* \times P_0] \times 0.0001 \tag{1.14}$$

with 0.0001 representing an interest rate change of one basis point (bp) or one hundredth of a percent. The **DVBP**, sometimes called the **DV01**, measures can be easily added up across the portfolio.

The second derivative is the **dollar convexity** (DC):

$$f''(y_0) = \frac{d^2P}{dy^2} = C \times P_0 \tag{1.15}$$

where C is called the **convexity**.

For fixed-income instruments with known cash flows, the price-yield function is known, and we can compute analytical first and second derivatives. Consider, for example, our simple zero-coupon bond in Equation (1.1) where the only payment is the face value, $C_T = F$. We take the first derivative, which is

$$\frac{dP}{dy} = \frac{d}{dy}\left[\frac{F}{(1+y)^T}\right] = (-T)\frac{F}{(1+y)^{T+1}} = -\frac{T}{(1+y)}P \tag{1.16}$$

Comparing with Equation (1.12), we see that the modified duration must be given by $D^* = T/(1+y)$. The conventional measure of **duration** is $D = T$, which does not include division by $(1+y)$ in the denominator. This is also called **Macaulay duration.** Note that duration is expressed in periods, like T. With annual compounding, duration is in years. With semiannual compounding, duration is in semesters. It then has to be divided by two for conversion to years. Modified

[3] Note that this chapter does not present duration in the traditional textbook order. In line with the advanced focus on risk management, we first analyze the properties of duration as a sensitivity measure. This applies to any type of fixed-income instrument. Later, we will illustrate the usual definition of duration as a weighted average maturity, which applies for fixed-coupon bonds only.

duration D^* is related to Macaulay duration D

$$D^* = \frac{D}{(1+y)} \tag{1.17}$$

Modified duration is the appropriate measure of interest rate exposure. The quantity $(1+y)$ appears in the denominator because we took the derivative of the present value term with discrete compounding. If we use continuous compounding, modified duration is identical to the conventional duration measure. In practice, the difference between Macaulay and modified duration is usually small.

Let us now go back to Equation (1.16) and consider the second derivative, which is

$$\frac{d^2P}{dy^2} = -(T+1)(-T)\frac{F}{(1+y)^{T+2}} = \frac{(T+1)T}{(1+y)^2} \times P \tag{1.18}$$

Comparing with Equation (1.15), we see that the convexity is $C = (T+1)T/(1+y)^2$. Note that its dimension is expressed in period squared. With semiannual compounding, convexity is measured in semesters squared. It then has to be divided by 4 for conversion to years squared.[4] So, convexity must be positive for bonds with fixed coupons.

Putting together all these equations, we get the Taylor expansion for the change in the price of a bond, which is

$$\Delta P = -[D^* \times P](\Delta y) + \frac{1}{2}[C \times P](\Delta y)^2 + \cdots \tag{1.19}$$

Therefore duration measures the first-order (linear) effect of changes in yield and convexity the second-order (quadratic) term.

Example: Computing the Price Approximation[5]

Consider a 10-year zero-coupon Treasury bond trading at a yield of 6 percent. The present value is obtained as $P = 100/(1+6/200)^{20} = 55.368$. As is the practice in the Treasury market, yields are semiannually compounded. Thus all computations should be carried out using semesters, after which final results can be converted into annual units.

Here, Macaulay duration is exactly 10 years, as $D = T$ for a zero coupon bond. Its modified duration is $D^* = 20/(1+6/200) = 19.42$ semesters, which is 9.71 years. Its convexity is $C = 21 \times 20/(1+6/200)^2 = 395.89$ semesters

[4] This is because the conversion to annual terms is obtained by multiplying the semiannual yield Δy by two. As a result, the duration term must be divided by 2 and the convexity term by 2^2, or 4, for conversion to annual units.

[5] For such examples in this handbook, please note that intermediate numbers are reported with fewer significant digits than actually used in the computations. As a result, using rounded off numbers may give results that differ slightly from the final numbers shown here.

squared, which is 98.97 in years squared. Dollar duration is $DD = D^* \times P = 9.71 \times \$55.37 = \$537.55$. The DVBP is $DVBP = DD \times 0.0001 = \$0.0538$.

We want to approximate the change in the value of the bond if the yield goes to 7%. Using Equation (1.19), we have $\Delta P = -[9.71 \times \$55.37](0.01) + 0.5[98.97 \times \$55.37](0.01)^2 = -\$5.375 + \$0.274 = -\$5.101$. Using the linear term only, the new price is $\$55.368 - \$5.375 = \$49.992$. Using the two terms in the expansion, the predicted price is slightly higher, at $\$55.368 - \$5.375 + \$0.274 = \$50.266$.

These numbers can be compared with the exact value, which is \$50.257. The linear approximation has a relative pricing error of −0.53%, which is not bad. Adding a quadratic term reduces this to an error of 0.02% only, which is very small, given typical bid-ask spreads.

More generally, Figure 1.2 compares the quality of the Taylor series approximation. We consider a 10-year bond paying a 6 percent coupon semiannually. Initially, the yield is also at 6 percent and, as a result, the price of the bond is at par, at \$100. The graph compares three lines representing

1. The actual, exact price $P = f(y_0 + \Delta y)$
2. The duration estimate $P = P_0 - D^* P_0 \Delta y$
3. The duration and convexity estimate $P = P_0 - D^* P_0 \Delta y + (1/2) C P_0 (\Delta y)^2$

The actual price curve shows an increase in the bond price if the yield falls and, conversely, a depreciation if the yield increases. This effect is captured by the tangent to the true price curve, which represents the linear approximation based on duration. For small movements in the yield, this linear approximation provides a reasonable fit to the exact price.

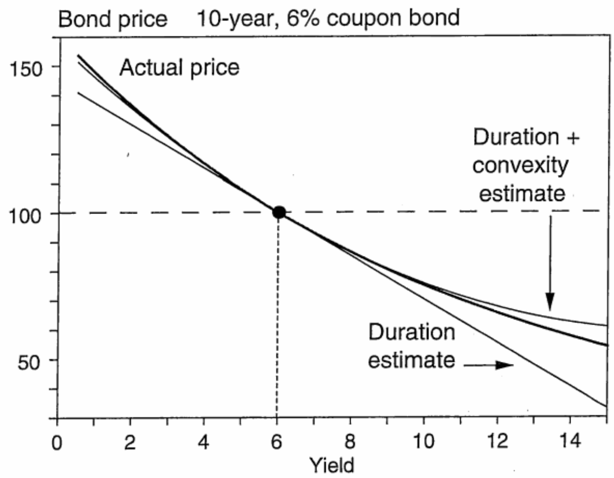

FIGURE 1.2 Price Approximation

KEY CONCEPT

Dollar duration measures the (negative) slope of the tangent to the price-yield curve at the starting point.

For large movements in price, however, the price-yield function becomes more curved and the linear fit deteriorates. Under these conditions, the quadratic approximation is noticeably better.

We should also note that the curvature is away from the origin, which explains the term *convexity* (as opposed to concavity). Figure 1.3 compares curves with different values for convexity. This curvature is beneficial since the second-order effect $0.5[C \times P](\Delta y)^2$ *must* be positive when convexity is positive.

As the figure shows, when the yield rises, the price drops but less than predicted by the tangent. Conversely, if the yield falls, the price increases faster than along the tangent. In other words, the quadratic term is always beneficial.

KEY CONCEPT

Convexity is always positive for regular coupon-paying bonds. Greater convexity is beneficial both for falling and rising yields.

The bond's modified duration and convexity can also be computed directly from numerical derivatives. Duration and convexity cannot be computed directly for some bonds, such as mortgage-backed securities, because their cash flows are uncertain. Instead, the portfolio manager has access to pricing models that can be used to reprice the securities under various yield environments.

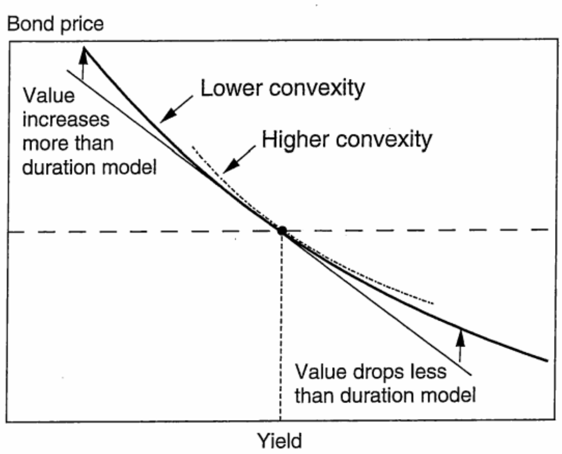

FIGURE 1.3 Effect of Convexity

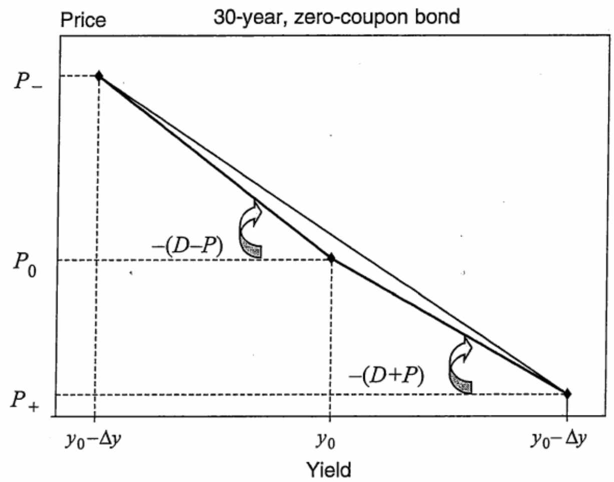

FIGURE 1.4 Effective Duration and Convexity

As shown in Figure 1.4, we choose a change in the yield, Δy, and reprice the bond under an upmove scenario, $P_+ = P(y_0 + \Delta y)$, and downmove scenario, $P_- = P(y_0 - \Delta y)$. **Effective duration** is measured by the numerical derivative. Using $D^* = -(1/P)dP/dy$, it is estimated as

$$D^E = \frac{[P_- - P_+]}{(2P_0\Delta y)} = \frac{P(y_0 - \Delta y) - P(y_0 + \Delta y)}{(2\Delta y)P_0} \tag{1.20}$$

Using $C = (1/P)d^2P/dy^2$, **effective convexity** is estimated as

$$C^E = [D_- - D_+]/\Delta y = \left[\frac{P(y_0 - \Delta y) - P_0}{(P_0\Delta y)} - \frac{P_0 - P(y_0 + \Delta y)}{(P_0\Delta y)}\right]/\Delta y \tag{1.21}$$

To illustrate, consider a 30-year zero-coupon bond with a yield of 6%, semi-annually compounded. The initial price is \$16.9733. We revalue the bond at 5% and 7%, with prices shown in the table. The effective duration in Equation (1.20) uses the two extreme points. The effective convexity in Equation (1.21) uses the difference between the dollar durations for the upmove and downmove. Note that convexity is positive if duration increases as yields fall, or if $D_- > D_+$.

The computations are detailed in Table 1.1, which shows an effective duration of 29.56. This is very close to the true value of 29.13, and would be even closer if the step Δy was smaller. Similarly, the effective convexity is 869.11, which is close to the true value of 862.48.

Finally, this numerical approach can be applied to get an estimate of the duration of a bond by considering bonds with the same maturity but different coupons. If interest rates decrease by 1%, the market price of a 6% bond should go up to a value close to that of a 7% bond. Thus we replace a drop in yield of Δy with an increase in coupon Δc and use the effective duration method to find

TABLE 1.1 Effective Duration and Convexity

State	Yield (%)	Bond Value	Duration Computation	Convexity Computation
Initial y_0	6.00	16.9733		
Up $y_0 + \Delta y$	7.00	12.6934		Duration up: 25.22
Down $y_0 - \Delta y$	5.00	22.7284		Duration down: 33.91
Difference in values			−10.0349	8.69
Difference in yields			0.02	0.01
Effective measure			29.56	869.11
Exact measure			29.13	862.48

the **coupon curve duration**[6]

$$D^{CC} = \frac{[P_+ - P_-]}{(2P_0\Delta c)} = \frac{P(y_0; c + \Delta c) - P(y_0; c - \Delta c)}{(2\Delta c)P_0} \tag{1.22}$$

This approach is useful for securities which are difficult to price under various yield scenarios. It only requires the market prices of securities with different coupons.

Example: Computation of Coupon Curve Duration

Consider a 10-year bond that pays a 7% coupon semiannually. In a 7% yield environment, the bond is selling at par and has modified duration of 7.11 years. The prices of 6% and 8% coupon bonds are \$92.89 and \$107.11, respectively. This gives a coupon curve duration of $(107.11 - 92.89)/(0.02 \times 100) = 7.11$, which in this case is the same as modified duration.

EXAMPLE 1.3: FRM EXAM 2006—QUESTION 75

A zero-coupon bond with a maturity of 10 years has an annual effective yield of 10%. What is the closest value for its modified duration?

a. 9
b. 10
c. 99
d. 100

[6] For a more formal proof, we could take the pricing formula for a consol at par and compute the derivatives with respect to y and c. Apart from the sign, these derivatives are identical when $y = c$.

EXAMPLE 1.4: FRM EXAM 2007—QUESTION 115

A portfolio manager has a bond position worth USD 100 million. The position has a modified duration of eight years and a convexity of 150 years. Assume that the term structure is flat. By how much does the value of the position change if interest rates increase by 25 basis points?

a. USD −2,046,875
b. USD −2,187,500
c. USD −1,953,125
d. USD −1,906,250

EXAMPLE 1.5: FRM EXAM 2007—QUESTION 55

Consider the following three methods of estimating the profit and loss (P&L) of a bullet bond: full repricing, duration (PV01), and duration plus convexity. Rank the methods to estimate the P&L impact of a large negative yield shock from the lowest to the highest.

a. Duration, duration plus convexity, full repricing
b. Duration, full repricing, duration plus convexity
c. Duration plus convexity, duration, full repricing
d. Full repricing, duration plus convexity, duration

1.3.1 Interpreting Duration and Convexity

The preceding section has shown how to compute analytical formulas for duration and convexity in the case of a simple zero-coupon bond. We can use the same approach for coupon-paying bonds. Going back to Equation (1.6), we have

$$\frac{dP}{dy} = \sum_{t=1}^{T} \frac{-tC_t}{(1+y)^{t+1}} = -[\sum_{t=1}^{T} \frac{tC_t}{(1+y)^t}]/P \times \frac{P}{(1+y)} = -\frac{D}{(1+y)}P \qquad (1.23)$$

which defines duration as

$$D = \sum_{t=1}^{T} \frac{tC_t}{(1+y)^t}/P \qquad (1.24)$$

The economic interpretation of duration is that it represents the average time to wait for each payment, weighted by the present value of the associated cash flow. Indeed, replacing P, we can write

$$D = \sum_{t=1}^{T} t \frac{C_t/(1+y)^t}{\sum C_t/(1+y)^t} = \sum_{t=1}^{T} t \times w_t \tag{1.25}$$

where the weights w_t represent the ratio of the present value of each cash flow C_t relative to the total, and sum to unity. This explains why the duration of a zero-coupon bond is equal to the maturity. There is only one cash flow and its weight is one.

KEY CONCEPT

(Macaulay) duration represents an average of the time to wait for all cash flows.

Figure 1.5 lays out the present value of the cash flows of a 6% coupon, 10-year bond. Given a duration of 7.80 years, this coupon-paying bond is equivalent to a zero-coupon bond maturing in exactly 7.80 years.

For bonds with fixed coupons, duration is less than maturity. For instance, Figure 1.6 shows how the duration of a 10-year bond varies with its coupon. With a zero coupon, Macaulay duration is equal to maturity. Higher coupons place more weight on prior payments and therefore reduce duration.

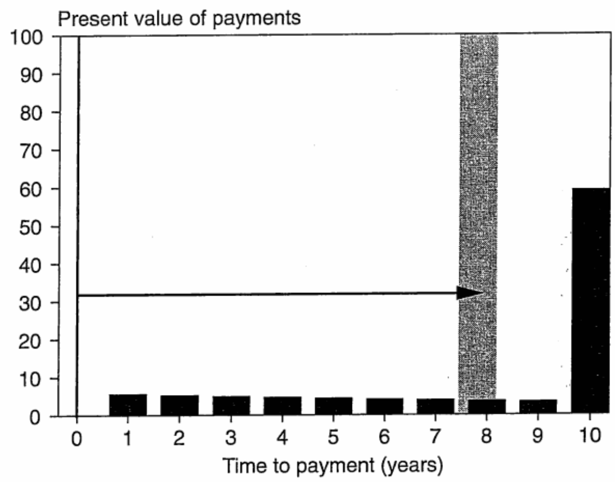

FIGURE 1.5 Duration as the Maturity of a Zero-Coupon Bond

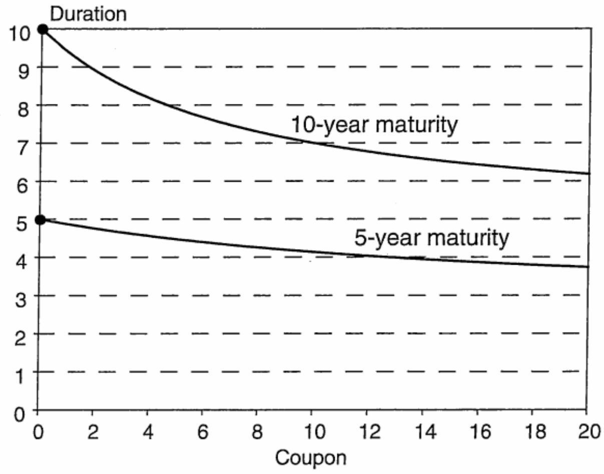

FIGURE 1.6 Duration and Coupon

Duration can be expressed in a simple form for **consols**. From Equation (1.8), we have $P = (c/y)F$. Taking the derivative, we find

$$\frac{dP}{dy} = cF\frac{(-1)}{y^2} = (-1)\frac{1}{y}[\frac{c}{y}F] = (-1)\frac{1}{y}P = -\frac{D_C}{(1+y)}P \tag{1.26}$$

Hence the Macaulay duration for the consol D_C is

$$D_C = \frac{(1+y)}{y} \tag{1.27}$$

This shows that the duration of a consol is finite even if its maturity is infinite. Also, this duration does not depend on the coupon.

This formula provides a useful rule of thumb. For a long-term coupon-paying bond, duration should be lower than $(1+y)/y$. For instance, when $y = 6\%$, the upper limit on duration is $D_C = 1.06/0.06$, or 17.7 years. In this environment, the duration of a par 30-year bond is 14.25, which is indeed lower than 17.7 years.

KEY CONCEPT

The duration of a long-term bond can be approximated by an upper bound, which is that of a consol with the same yield, $D_C = (1+y)/y$.

Figure 1.7 describes the relationship between duration, maturity, and coupon for regular bonds in a 6% yield environment. For the zero-coupon bond, $D = T$, which is a straight line going through the origin. For the par 6% bond, duration increases monotonically with maturity until it reaches the asymptote of D_C. The

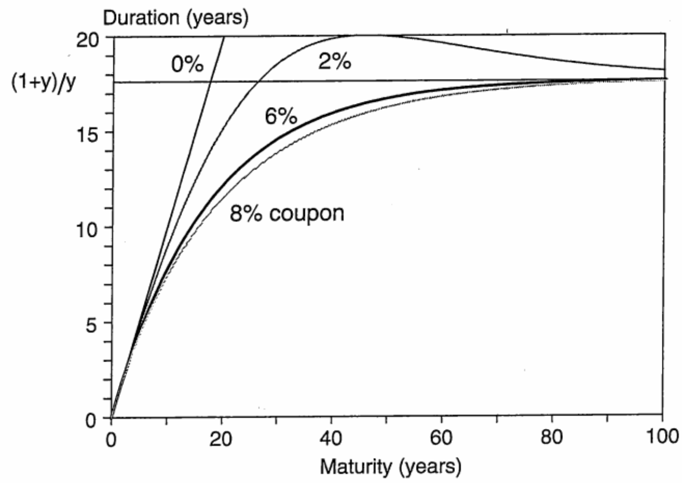

FIGURE 1.7 Duration and Maturity

8% bond has lower duration than the 6% bond for fixed T. Greater coupons, for a fixed maturity, decrease duration, as more of the payments come early.

Finally, the 2% bond displays a pattern intermediate between the zero-coupon and 6% bonds. It initially behaves like the zero, exceeding D_C initially then falling back to the asymptote, which is the same for all coupon-paying bonds.

Taking now the second derivative in Equation (1.23), we have

$$\frac{d^2P}{dy^2} = \sum_{t=1}^{T} \frac{t(t+1)C_t}{(1+y)^{t+2}} = \left[\sum_{t=1}^{T} \frac{t(t+1)C_t}{(1+y)^{t+2}} / P \right] \times P \tag{1.28}$$

which defines convexity as

$$C = \sum_{t=1}^{T} \frac{t(t+1)C_t}{(1+y)^{t+2}} / P \tag{1.29}$$

Convexity can also be written as

$$C = \sum_{t=1}^{T} \frac{t(t+1)}{(1+y)^2} \times \frac{C_t/(1+y)^t}{\sum C_t/(1+y)^t} = \sum_{t=1}^{T} \frac{t(t+1)}{(1+y)^2} \times w_t \tag{1.30}$$

Because the squared t term dominates in the fraction, this basically involves a weighted average of the square of time. Therefore, convexity is much greater for long-maturity bonds because they have payoffs associated with large values of t. The formula also shows that convexity is always positive for such bonds, implying that the curvature effect is beneficial. As we will see later, convexity can be negative for bonds that have uncertain cash flows, such as **mortgage-backed securities** (MBSs) or callable bonds.

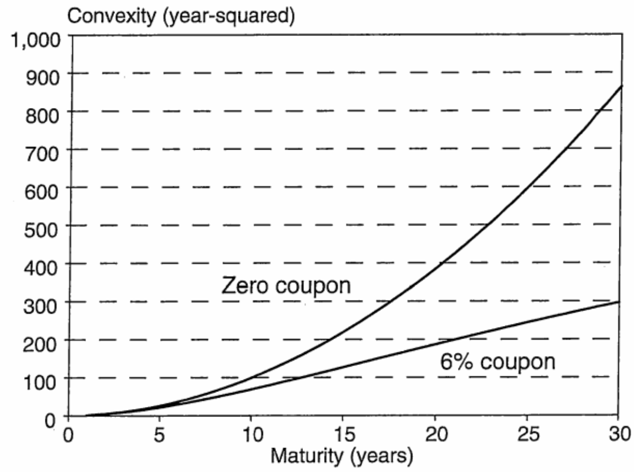

FIGURE 1.8 Convexity and Maturity

Figure 1.8 displays the behavior of convexity, comparing a zero-coupon bond with a 6% coupon bond with identical maturities. The zero-coupon bond always has greater convexity, because there is only one cash flow at maturity. Its convexity is roughly the square of maturity, for example about 900 for the 30-year zero. In contrast, the 30-year coupon bond has a convexity of about 300 only.

KEY CONCEPT

All else equal, duration and convexity both increase for longer maturities, lower coupons, and lower yields.

As an illustration, Table 1.2 details the steps of the computation of duration and convexity for a two-year, 6% semiannual coupon-paying bond. We first convert the annual coupon and yield into semiannual equivalent, \$3 and 3% each. The PV column then reports the present value of each cash flow. We verify that these add up to \$100, since the bond must be selling at par.

Next, the duration term column multiplies each PV term by time, or more precisely the number of half years until payment. This adds up to \$382.86, which divided by the price gives $D = 3.83$. This number is measured in half years, and we need to divide by two to convert to years. Macaulay duration is 1.91 years, and modified duration $D^* = 1.91/1.03 = 1.86$ years. Note that, to be consistent, the adjustment in the denominator involves the semiannual yield of 3%.

Finally, the right-most column shows how to compute the bond's convexity. Each term involves PV_t times $t(t+1)/(1+y)^2$. These terms sum to 1,777.755, or divided by the price, 17.78. This number is expressed in units of time squared and must be divided by 4 to be converted in annual terms. We find a convexity of $C = 4.44$, in year-squared.

TABLE 1.2 Computing Duration and Convexity

Period (half-year) t	Payment C_t	Yield (%) (6 mo)	PV of Payment $C_t/(1+y)^t$	Duration Term tPV_t	Convexity Term $t(t+1)PV_t \times (1/(1+y)^2)$
1	3	3.00	2.913	2.913	5.491
2	3	3.00	2.828	5.656	15.993
3	3	3.00	2.745	8.236	31.054
4	103	3.00	91.514	366.057	1725.218
Sum:			100.00	382.861	1777.755
(half-years)				3.83	17.78
(years)				1.91	
Modified duration				1.86	
Convexity					4.44

EXAMPLE 1.6: FRM EXAM 2003—QUESTION 13

Suppose the face value of a three-year option-free bond is USD 1,000 and the annual coupon is 10%. The current yield to maturity is 5%. What is the modified duration of this bond?

a. 2.62
b. 2.85
c. 3.00
d. 2.75

EXAMPLE 1.7: FRM EXAM 2002—QUESTION 118

A Treasury bond has a coupon rate of 6% per annum (the coupons are paid semiannually) and a semiannually compounded yield of 4% per annum. The bond matures in 18 months and the next coupon will be paid 6 months from now. Which number below is closest to the bond's Macaulay duration?

a. 1.023 years
b. 1.457 years
c. 1.500 years
d. 2.915 years

EXAMPLE 1.8: DURATION AND COUPON

A and B are two perpetual bonds, that is, their maturities are infinite. A has a coupon of 4% and B has a coupon of 8%. Assuming that both are trading at the same yield, what can be said about the duration of these bonds?

a. The duration of A is greater than the duration of B.
b. The duration of A is less than the duration of B.
c. A and B both have the same duration.
d. None of the above.

EXAMPLE 1.9: FRM EXAM 2004—QUESTION 16

A manager wants to swap a bond for a bond with the same price but higher duration. Which of the following bond characteristics would be associated with a higher duration?

I. A higher coupon rate
II. More frequent coupon payments
III. A longer term to maturity
IV. A lower yield

a. I, II, and III
b. II, III, and IV
c. III and IV
d. I and II

EXAMPLE 1.10: FRM EXAM 2001—QUESTION 104

When the maturity of a plain coupon bond increases, its duration increases

a. Indefinitely and regularly
b. Up to a certain level
c. Indefinitely and progressively
d. In a way dependent on the bond being priced above or below par

EXAMPLE 1.11: FRM EXAM 2000—QUESTION 106

Consider the following bonds:

Bond Number	Maturity (yrs)	Coupon Rate	Frequency	Yield (Annual)
1	10	6%	1	6%
2	10	6%	2	6%
3	10	0%	1	6%
4	10	6%	1	5%
5	9	6%	1	6%

How would you rank the bonds from the shortest to longest duration?

- **a.** 5-2-1-4-3
- **b.** 1-2-3-4-5
- **c.** 5-4-3-1-2
- **d.** 2-4-5-1-3

EXAMPLE 1.12: FRM EXAM 2000—QUESTION 110

Which of the following statements are *true*?

I. The convexity of a 10-year zero-coupon bond is higher than the convexity of a 10-year, 6% bond.
II. The convexity of a 10-year zero-coupon bond is higher than the convexity of a 6% bond with a duration of 10 years.
III. Convexity grows proportionately with the maturity of the bond.
IV. Convexity is always positive for all types of bonds.
V. Convexity is always positive for "straight" bonds.

- **a.** I only
- **b.** I and II only
- **c.** I and V only
- **d.** II, III, and V only

1.3.2 Portfolio Duration and Convexity

Fixed-income portfolios often involve very large numbers of securities. It would be impractical to consider the movements of each security individually. Instead, portfolio managers aggregate the duration and convexity across the portfolio. A manager who believes that rates will increase should shorten the portfolio duration relative to that of the benchmark. Say for instance that the benchmark has a duration of 5 years. The manager shortens the portfolio duration to 1 year only. If rates increase by 2%, the benchmark will lose approximately $5y \times 2\% = 10\%$. The portfolio, however, will only lose $1y \times 2\% = 2\%$, hence "beating" the benchmark by 8%.

Because the Taylor expansion involves a summation, the portfolio duration is easily obtained from the individual components. Say we have N components indexed by i. Defining D_p^* and P_p as the portfolio modified duration and value, the portfolio dollar duration (DD) is

$$D_p^* P_p = \sum_{i=1}^{N} D_i^* x_i P_i \tag{1.31}$$

where x_i is the number of units of bond i in the portfolio. A similar relationship holds for the portfolio dollar convexity (DC). If yields are the same for all components, this equation also holds for the Macaulay duration.

Because the portfolio's total market value is simply the summation of the component market values,

$$P_p = \sum_{i=1}^{N} x_i P_i \tag{1.32}$$

we can define the **portfolio weight** w_i as $w_i = x_i P_i / P_p$, provided that the portfolio market value is nonzero. We can then write the portfolio duration as a weighted average of individual durations

$$D_p^* = \sum_{i=1}^{N} D_i^* w_i \tag{1.33}$$

Similarly, the portfolio convexity is a weighted average of convexity numbers

$$C_p = \sum_{i=1}^{N} C_i w_i \tag{1.34}$$

As an example, consider a portfolio invested in three bonds, described in Table 1.3. The portfolio is long a 10-year and 1-year bond, and short a 30-year zero-coupon bond. Its market value is \$1,301,600. Summing the duration for each component, the portfolio dollar duration is \$2,953,800, which translates into a

TABLE 1.3 Portfolio Dollar Duration and Convexity

	Bond 1	Bond 2	Bond 3	Portfolio
Maturity (years)	10	1	30	
Coupon	6%	0%	0%	
Yield	6%	6%	6%	
Price P_i	\$100.00	\$94.26	\$16.97	
Modified duration D_i^*	7.44	0.97	29.13	
Convexity C_i	68.78	1.41	862.48	
Number of bonds x_i	10,000	5,000	−10,000	
Dollar amounts $x_i P_i$	\$1,000,000	\$471,300	−\$169,700	\$1,301,600
Weight w_i	76.83%	36.21%	−13.04%	100.00%
Dollar duration $D_i^* P_i$	\$744.00	\$91.43	\$494.34	
Portfolio DD: $x_i D_i^* P_i$	\$7,440,000	\$457,161	−\$4,943,361	\$2,953,800
Portfolio DC: $x_i C_i P_i$	68,780,000	664,533	−146,362,856	−76,918,323

duration of 2.27 years. The portfolio convexity is $-76{,}918{,}323/1,301,600 = -59.10$, which is negative due to the short position in the 30-year zero, which has very high convexity.

Alternatively, assume the portfolio manager is given a benchmark which is the first bond. He or she wants to invest in bonds 2 and 3, keeping the portfolio duration equal to that of the target, or 7.44 years. To achieve the target value and dollar duration, the manager needs to solve a system of two equations in the numbers x_1 and x_2:

$$\begin{array}{lrrr} \text{Value:} & \$100 = & x_1\$94.26 + & x_2\$16.97 \\ \text{Dol.Duration:} & 7.44 \times \$100 = & 0.97 \times x_1\$94.26 + & 29.13 \times x_2\$16.97 \end{array}$$

The solution is $x_1 = 0.817$ and $x_2 = 1.354$, which gives a portfolio value of \$100 and modified duration of 7.44 years.[7] The portfolio convexity is 199.25, higher than the index. Such a portfolio consisting of very short and very long maturities is called a **barbell portfolio.** In contrast, a portfolio with maturities in the same range is called a **bullet portfolio.** Note that the barbell portfolio has a much greater convexity than the bullet bond because of the payment in 30 years. Such a portfolio would be expected to outperform the bullet portfolio if yields moved by a large amount.

In sum, duration and convexity are key measures of fixed-income portfolios. They summarize the linear and quadratic exposure to movements in yields. This explains why they are essential tools for fixed-income portfolio managers.

[7] This can be obtained by first expressing x_2 in the first equation as a function of x_1 and then substituting back into the second equation. This gives $x_2 = (100 - 94.26x_1)/16.97$, and $744 = 91.43x_1 + 494.34x_2 = 91.43x_1 + 494.34(100 - 94.26x_1)/16.97 = 91.43x_1 + 2913.00 - 2745.79x_1$. Solving, we find $x_1 = (-2169.00)/(-2654.36) = 0.817$ and $x_2 = (100 - 94.26 \times 0.817)/16.97 = 1.354$.

EXAMPLE 1.13: FRM EXAM 2002—QUESTION 57

A bond portfolio has the following composition:

1. Portfolio A: price $90,000, modified duration 2.5, long position in 8 bonds
2. Portfolio B: price $110,000, modified duration 3, short position in 6 bonds
3. Portfolio C: price $120,000, modified duration 3.3, long position in 12 bonds

All interest rates are 10%. If the rates rise by 25 basis points, then the bond portfolio value will

a. Decrease by $11,430
b. Decrease by $21,330
c. Decrease by $12,573
d. Decrease by $23,463

EXAMPLE 1.14: FRM EXAM 2006—QUESTION 61

Consider the following portfolio of bonds (par amounts are in millions of USD).

Bond	Price	Par amount held	Modified Duration
A	101.43	3	2.36
B	84.89	5	4.13
C	121.87	8	6.27

What is the value of the portfolio's DV01 (dollar value of 1 basis point)?

a. 8,019
b. 8,294
c. 8,584
d. 8,813

1.4 IMPORTANT FORMULAS

Compounding: $(1+y)^T = (1+y^S/2)^{2T} = e^{y^C T}$

Fixed-coupon bond valuation: $P = \sum_{t=1}^{T} \frac{C_t}{(1+y)^t}$

Taylor expansion: $P_1 = P_0 + f'(y_0)\Delta y + \frac{1}{2} f''(y_0)(\Delta y)^2 + \cdots$

Duration as exposure: $\frac{dP}{dy} = -D^* \times P$, $\text{DD} = D^* \times P$, $\text{DVBP} = \text{DD} \times 0.0001$

Conventional duration: $D^* = \frac{D}{(1+y)}$, $D = \sum_{t=1}^{T} \frac{tC_t}{(1+y)^t}/P$

Convexity: $\frac{d^2P}{dy^2} = C \times P$, $C = \sum_{t=1}^{T} \frac{t(t+1)C_t}{(1+y)^{t+2}}/P$

Price change: $\Delta P = -[D^* \times P](\Delta y) + 0.5[C \times P](\Delta y)^2 + \cdots$

Consol: $P = \frac{c}{y}F$, $D = \frac{(1+y)}{y}$

Portfolio duration and convexity: $D_p^* = \sum_{i=1}^{N} D_i^* w_i$, $C_p = \sum_{i=1}^{N} C_i w_i$

1.5 ANSWERS TO CHAPTER EXAMPLES

Example 1.1: FRM Exam 2002—Question 48

a. The EAR is defined by $FV/PV = (1 + \text{EAR})^T$. So $\text{EAR} = (FV/PV)^{1/T} - 1$. Here, $T = 1/12$. So, $\text{EAR} = (1{,}000/987)^{12} - 1 = 17.0\%$.

Example 1.2: FRM Exam 2002—Question 51

c. The time T relates the current and future values such that $FV/PV = 2 = (1 + 8\%)^T$. Taking logs of both sides, this gives $T = \ln(2)/\ln(1.08) = 9.006$.

Example 1.3: FRM Exam 2006—Question 75

a. Without doing any computation, the Macaulay duration must be 10 years because this is a zero-coupon bond. With annual compounding, modified duration is $D^* = 10/(1 + 10\%)$, or close to 9 years.

Example 1.4: FRM Exam 2007—Question 115

c. The change in price is given by $\Delta P = -[D^* \times P](\Delta y) + \frac{1}{2}[C \times P](\Delta y)^2$ $= -[8 \times 100](0.0025) + 0.5[150 \times 100](0.0025)^2 = -2.000000 + 0.046875 = -1.953125$.

Example 1.5: FRM Exam 2007—Question 55

a. When yields drop, the duration approximation gives the smallest price increase, so the answer must be either a. or b. Figure 1.2 shows that the full repricing curve for decreases in yields is slightly higher than the duration and convexity approximation. Alternatively, differentiating Equation 1.18 once more give a negative term for the third-order derivative. Combined with δy^3, which is negative, the third-order term must be positive.

Example 1.6: FRM Exam 2003—Question 13

d. As in Table 1.2, we lay out the cash flows and find

Period t	Payment C_t	Yield y	$PV_t = C_t/(1+y)^t$	tPV_t
1	100	5.00	95.24	95.24
2	100	5.00	90.71	181.41
3	1100	5.00	950.22	2850.66
Sum:			1136.16	3127.31

Duration is then 2.75, and modified duration 2.62.

Example 1.7: FRM Exam 2002—Question 118

b. For coupon-paying bonds, Macaulay duration is slightly less than the maturity, which is 1.5 year here. So, b. would be a good guess. Otherwise, we can compute duration exactly.

Example 1.8: Duration and Coupon

c. Going back to the duration equation for the consol, Equation (1.27), we see that it does not depend on the coupon but only on the yield. Hence, the durations must be the same. The price of bond A, however, must be half that of bond B.

Example 1.9: FRM Exam 2004—Question 16

c. Higher duration is associated with physical characteristics that push payments into the future, i.e., longer term, lower coupons, and less frequent coupon payments, as well as lower yields, which increase the relative weight of payments in the future.

Example 1.10: FRM Exam 2001—Question 104

b. With a fixed coupon, the duration goes up to the level of a consol with the same coupon. See Figure 1.7.

Example 1.11: FRM Exam 2000—Question 106

a. The nine-year bond (number 5) has shorter duration because the maturity is shortest, at nine years, among comparable bonds. Next, we have to decide between bonds 1 and 2, which only differ in the payment frequency. The semiannual bond (number 2) has a first payment in six months and has shorter duration than the annual bond. Next, we have to decide between bonds 1 and 4, which only differ in the yield. With lower yield, the cash flows further in the future have a higher weight, so that bond 4 has greater duration. Finally, the zero-coupon bond has the longest duration. So, the order is 5-2-1-4-3.

Example 1.12: FRM Exam 2000—Question 110

c. Because convexity is proportional to the square of time to payment, the convexity of a bond is mainly driven by the cash flows far into the future. Answer I. is correct because the 10-year zero has only one cash flow, whereas the coupon bond has several others that reduce convexity. Answer II. is false because the 6% bond with 10-year duration must have cash flows much further into the future, say in 30 years, which will create greater convexity. Answer III. is false because convexity grows with the square of time. Answer IV. is false because some bonds, for example MBSs or callable bonds, can have negative convexity. Answer V. is correct because convexity must be positive for coupon-paying bonds.

Example 1.13: FRM Exam 2002—Question 57

a. The portfolio dollar duration is $D^*P = \sum x_i D_i^* P_i = +8 \times 2.5 \times \$90{,}000 - 6 \times 3.0 \times \$110{,}000 + 12 \times 3.3 \times \$120{,}000 = \$4{,}572{,}000$. The change in portfolio value is then $-(D * P)(\Delta y) = -\$4{,}572{,}000 \times 0.0025 = -\$11{,}430$.

Example 1.14: FRM Exam 2006—Question 61

c. First, the market value of each bond is obtained by multiplying the par amount by the ratio of the market price divided by 100. Next, this is multiplied by D^* to get the dollar duration DD. Summing, this gives \$85.841 million. We multiply by 1,000,000 to get dollar amounts and by 0.0001 to get the DV01, which gives \$8,584.

Bond	Price	Par	Mkt value	D^*	DD
A	101.43	3	3.043	2.36	7.181
B	84.89	5	4.245	4.13	15.530
C	121.87	8	9.750	6.27	61.130
Sum		8			85.841

APPENDIX: APPLICATIONS OF INFINITE SERIES

When bonds have fixed coupons, the bond valuation problem often can be interpreted in terms of combinations of infinite series. The most important infinite series result is for a sum of terms that increase at a geometric rate:

$$1 + a + a^2 + a^3 + \cdots = \frac{1}{1-a} \tag{1.35}$$

This can be proved, for instance, by multiplying both sides by $(1 - a)$ and canceling out terms.

Equally important, consider a geometric series with a finite number of terms, say N. We can write this as the difference between two infinite series:

$$1 + a + a^2 + a^3 + \cdots + a^{N-1}$$
$$= (1 + a + a^2 + a^3 + \cdots) - a^N(1 + a + a^2 + a^3 + \cdots) \qquad (1.36)$$

such that all terms with order N or higher will cancel each other.

We can then write

$$1 + a + a^2 + a^3 + \cdots + a^{N-1} = \frac{1}{1-a} - a^N \frac{1}{1-a} \qquad (1.37)$$

These formulas are essential to value bonds. Consider first a consol with an infinite number of coupon payments with a fixed coupon rate c. If the yield is y and the face value F, the value of the bond is

$$\begin{aligned}
P &= cF\left[\tfrac{1}{(1+y)} + \tfrac{1}{(1+y)^2} + \tfrac{1}{(1+y)^3} + \cdots\right] \\
&= cF\tfrac{1}{(1+y)}[1 + a^2 + a^3 + \cdots] \\
&= cF\tfrac{1}{(1+y)}\left[\tfrac{1}{1-a}\right] \\
&= cF\tfrac{1}{(1+y)}\left[\tfrac{1}{(1-1/(1+y))}\right] \\
&= cF\tfrac{1}{(1+y)}\left[\tfrac{(1+y)}{y}\right] \\
&= \tfrac{c}{y}F
\end{aligned}$$

Similarly, we can value a bond with a *finite* number of coupons over T periods at which time the principal is repaid. This is really a portfolio with three parts:

1. A long position in a consol with coupon rate c
2. A short position in a consol with coupon rate c that starts in T periods
3. A long position in a zero-coupon bond that pays F in T periods.

Note that the combination of (1) and (2) ensures that we have a finite number of coupons. Hence, the bond price should be:

$$P = \frac{c}{y}F - \frac{1}{(1+y)^T}\frac{c}{y}F + \frac{1}{(1+y)^T}F = \frac{c}{y}F\left[1 - \frac{1}{(1+y)^T}\right] + \frac{1}{(1+y)^T}F \qquad (1.38)$$

where again the formula can be adjusted for different compounding methods.

This is useful for a number of purposes. For instance, when $c = y$, it is immediately obvious that the price must be at par, $P = F$. This formula also can be used to find closed-form solutions for duration and convexity.

CHAPTER 2

Fundamentals of Probability

The preceding chapter has laid out the foundations for understanding how bond prices move in relation to yields. More generally, the instrument can be described by a payoff function, which links the price to the underlying risk factor. Next, we have to characterize movements in bond yields, or more generally, any relevant risk factor in financial markets.

This is done with the tools of probability, a mathematical abstraction that describes the distribution of risk factors. Each risk factor is viewed as a random variable whose properties are described by a probability distribution function. These distributions can be processed with the payoff function to create a distribution of the profit and loss profile for the trading portfolio.

This chapter reviews the fundamental tools of probability theory for risk managers. Section 2.1 lays out the foundations, characterizing random variables by their probability density and distribution functions. These functions can be described by their principal moments, mean, variance, skewness, and kurtosis. Distributions with multiple variables are described in Section 2.2. Section 2.3 then turns to functions of random variables. Section 2.4 presents some examples of important distribution functions for risk management, including the uniform, normal, lognormal, Student's, binomial, and Poisson. Finally, Section 2.5 discusses limit distributions, which can be used to characterize the average and tails of independent random variables.

2.1 CHARACTERIZING RANDOM VARIABLES

The classical approach to probability is based on the concept of the **random variable** (rv). This can be viewed as the outcome from throwing a die, for example. Each realization is generated from a fixed process. If the die is perfectly symmetrical, with six faces, we could say that the probability of observing a face with a six in one throw is $p = 1/6$. Although the event itself is random, we can still make a number of useful statements from a fixed data-generating process.

The same approach can be taken to financial markets, where stock prices, exchange rates, yields, and commodity prices can be viewed as random variables. The assumption of a fixed data-generating process for these variables, however, is more tenuous than for the preceding experiment.

2.1.1 Univariate Distribution Functions

A random variable X is characterized by a **distribution function,**

$$F(x) = P(X \leq x) \tag{2.1}$$

which is the probability that the realization of the random variable X ends up less than or equal to the given number x. This is also called a **cumulative distribution function.**

When the variable X takes discrete values, this distribution is obtained by summing the step values less than or equal to x. That is,

$$F(x) = \sum_{x_j \leq x} f(x_j) \tag{2.2}$$

where the function $f(x)$ is called the **frequency function** or the **probability density function** (p.d.f.). Here, $f(x)$ is the probability of observing x. This function is characterized by its shape as well as fixed parameters, θ.

When the variable is continuous, the distribution is given by

$$F(x) = \int_{-\infty}^{x} f(u)du \tag{2.3}$$

The density can be obtained from the distribution using

$$f(x) = \frac{dF(x)}{dx} \tag{2.4}$$

Often, the random variable will be described interchangeably by its distribution or its density.

These functions have notable properties. The density $f(u)$ must be positive for all u. As x tends to infinity, the distribution tends to unity as it represents the total probability of any draw for x:

$$\int_{-\infty}^{\infty} f(u)du = 1 \tag{2.5}$$

Figure 2.1 gives an example of a density function $f(x)$, on the top panel, and of a cumulative distribution function $F(x)$ on the bottom panel. $F(x)$ measures the area under the $f(x)$ curve to the left of x, which is represented by the shaded area. Here, this area is 0.24. For small values of x, $F(x)$ is close to zero. Conversely, for large values of x, $F(x)$ is close to unity.

Example: Density Functions

A gambler wants to characterize the probability density function of the outcomes from a pair of dice. Because each has six faces, there are $6^2 = 36$ possible throw combinations. Out of these, there is one occurrence of an outcome of two (each die showing one). So, the frequency of an outcome of two is one. We can have two occurrences of a three (a one and a two and vice versa), and so on.

The gambler compiles the frequency of each value, from 2 to 12, as shown in Table 2.1. From this, he or she can compute the probability of each outcome. For instance, the probability of observing three is equal to 2, the frequency $n(x)$, divided by the total number of outcomes, of 36, which gives 0.0556. We can verify that all the probabilities indeed add up to one, since all occurrences must be accounted for. From the table, we see that the probability of an outcome of 3 or less is 8.33%.

TABLE 2.1 Probability Density Function

Outcome x_i	Frequency $n(x)$	Probability $f(x)$		Cumulative Probability $F(x)$
2	1	1/36	0.0278	0.0278
3	2	2/36	0.0556	0.0833
4	3	3/36	0.0833	0.1667
5	4	4/36	0.1111	0.2778
6	5	5/36	0.1389	0.4167
7	6	6/36	0.1667	0.5833
8	5	5/36	0.1389	0.7222
9	4	4/36	0.1111	0.8333
10	3	3/36	0.0833	0.9167
11	2	2/36	0.0556	0.9722
12	1	1/36	0.0278	1.0000
Sum	36	1	1.0000	

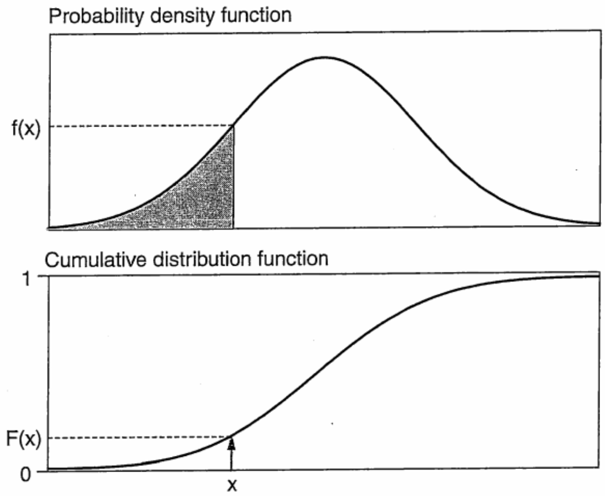

FIGURE 2.1 Density and Distribution Functions

2.1.2 Moments

A random variable is characterized by its distribution function. Instead of having to report the whole function, it is convenient to summarize it by a few parameters, or **moments**.

For instance, the expected value for x, or **mean**, is given by the integral

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx \tag{2.6}$$

which measures the *central tendency*, or *center of gravity* of the population.

The distribution can also be described by its **quantile**, which is the cutoff point x with an associated probability c:

$$F(x) = \int_{-\infty}^{x} f(u) du = c \tag{2.7}$$

So, there is a probability of c that the random variable will fall *below* x. Because the total probability adds up to one, there is a probability of $p = 1 - c$ that the random variable will fall *above* x. Define this quantile as $Q(X, c)$. The 50% quantile is known as the **median**.

In fact, value at risk (VAR) can be interpreted as the cutoff point such that a loss will not happen with probability greater than $p = 95\%$, say. If $f(u)$ is the distribution of profit and losses on the portfolio, VAR is defined from

$$F(x) = \int_{-\infty}^{x} f(u) du = (1 - p) \tag{2.8}$$

where p is the right-tail probability, and c the usual left-tail probability. VAR can be defined as minus the quantile itself, or alternatively, the deviation between the expected value and the quantile,

$$\mathrm{VAR}(c) = E(X) - Q(X, c) \tag{2.9}$$

Note that VAR is typically reported as a loss, i.e., a positive number, which explains the negative sign. Figure 2.2 shows an example with $c = 5\%$.

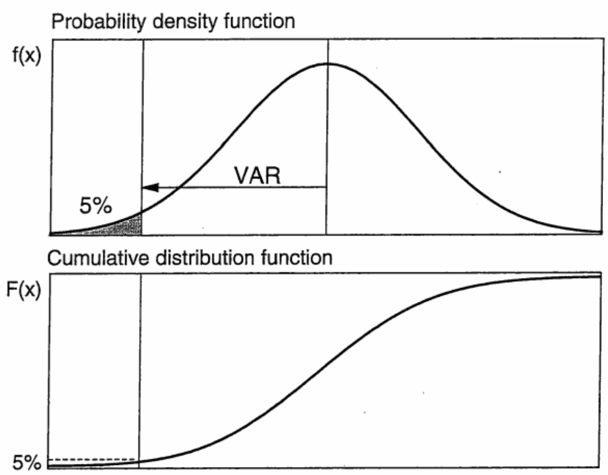

FIGURE 2.2 VAR as a Quantile

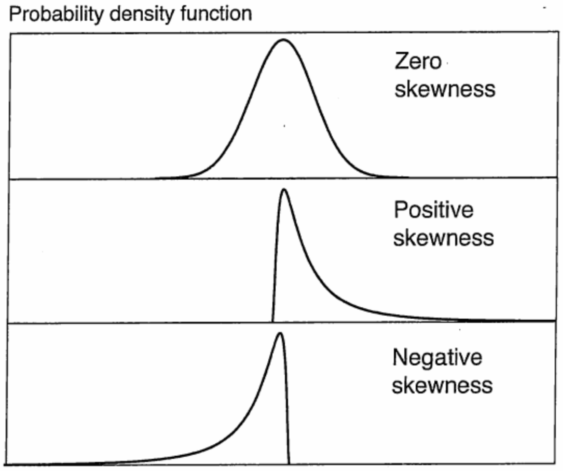

FIGURE 2.3 Effect of Skewness

Another useful moment is the squared dispersion around the mean, or **variance**

$$\sigma^2 = V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx \tag{2.10}$$

The **standard deviation** is more convenient to use as it has the same units as the original variable X

$$\text{SD}(X) = \sigma = \sqrt{V(X)} \tag{2.11}$$

Next, the scaled third moment is the **skewness**, which describes departures from symmetry. It is defined as

$$\gamma = \left(\int_{-\infty}^{+\infty} [x - E(X)]^3 f(x)dx \right) \Big/ \sigma^3 \tag{2.12}$$

Negative skewness indicates that the distribution has a long left tail, which indicates a high probability of observing large negative values. If this represents the distribution of profits and losses for a portfolio, this is a dangerous situation. Figure 2.3 displays distributions with various signs for the skewness.

The scaled fourth moment is the **kurtosis**, which describes the degree of "flatness" of a distribution, or width of its tails. It is defined as

$$\delta = \left(\int_{-\infty}^{+\infty} [x - E(X)]^4 f(x)dx \right) \Big/ \sigma^4 \tag{2.13}$$

Because of the fourth power, large observations in the tail will have a large weight and hence create large kurtosis. Such a distribution is called **leptokurtic**, or **fat-tailed**. This parameter is very important for risk measurement. A kurtosis of 3 is considered average. High kurtosis indicates a higher probability of extreme movements. A distribution with kurtosis lower than 3 is called **platykurtic**. Figure 2.4 displays distributions with various values for the kurtosis.

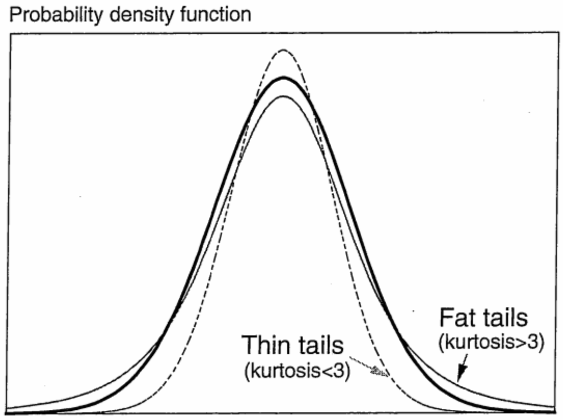

FIGURE 2.4 Effect of Kurtosis

Example: Computing Moments

Our gambler wants to know the expected value of the outcome of throwing two dice. He computes the product of each outcome and associated probability, as shown in Table 2.2. For instance, the first entry is $xf(x) = 2 \times 0.0278 = 0.0556$, and so on. Summing across all events, the mean is $\mu = 7.000$. This is also the median, since the distribution is perfectly symmetrical.

Next, we can use Equation (2.10) to compute the variance. The first term is $(x-\mu)^2\ f(x) = (2-7)^2 0.0278 = 0.6944$. These terms add up to 5.8333, or, taking the square root, $\sigma = 2.4152$. The skewness terms sum to zero, because for each entry with a positive deviation $(x-\mu)^3$, there is an identical one with a negative sign and with the same probability. Finally, the kurtosis terms $(x-\mu)^4 f(x)$ sum to 80.5. Dividing by $\sigma^4 = 34.0278$, this gives a kurtosis of $\delta = 2.3657$.

TABLE 2.2 Computing Moments of a Distribution

Outcome x_i	Prob. $f(x)$	Mean $xf(x)$	Variance $(x-\mu)^2 f(x)$	Skewness $(x-\mu)^3 f(x)$	Kurtosis $(x-\mu)^4 f(x)$
2	0.0278	0.0556	0.6944	−3.4722	17.3611
3	0.0556	0.1667	0.8889	−3.5556	14.2222
4	0.0833	0.3333	0.7500	−2.2500	6.7500
5	0.1111	0.5556	0.4444	−0.8889	1.7778
6	0.1389	0.8333	0.1389	−0.1389	0.1389
7	0.1667	1.1667	0.0000	0.0000	0.0000
8	0.1389	1.1111	0.1389	0.1389	0.1389
9	0.1111	1.0000	0.4444	0.8889	1.7778
10	0.0833	0.8333	0.7500	2.2500	6.7500
11	0.0556	0.6111	0.8889	3.5556	14.2222
12	0.0278	0.3333	0.6944	3.4722	17.3611
Sum	1.0000	7.0000	$\sigma^2 = 5.8333$	0.0000	80.5000
Denominator				$\sigma^3 = 14.0888$	$\sigma^4 = 34.0278$
		Mean	StdDev	Skewness	Kurtosis
		$\mu = 7.00$	$\sigma = 2.4152$	$\gamma = 0.0000$	$\delta = 2.3657$

2.2 MULTIVARIATE DISTRIBUTION FUNCTIONS

In practice, portfolio payoffs depend on numerous random variables. To simplify, start with two random variables. This could represent two currencies, or two interest rate factors, or default and credit exposure, to give just a few examples.

2.2.1 Joint Distributions

We can extend Equation (2.1) to

$$F_{12}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) \tag{2.14}$$

which defines a joint bivariate distribution function. In the continuous case, this is also

$$F_{12}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{12}(u_1, u_2) du_1 du_2 \tag{2.15}$$

where $f(u_1, u_2)$ is now the **joint density**. In general, adding random variables considerably complicates the characterization of the density or distribution functions.

The analysis simplifies considerably if the variables are **independent**. In this case, the joint density separates out into the product of the densities:

$$f_{12}(u_1 u_2) = f_1(u_1) \times f_2(u_2) \tag{2.16}$$

and the integral reduces to

$$F_{12}(x_1, x_2) = F_1(x_1) \times F_2(x_2) \tag{2.17}$$

This is very convenient because we only need to know the individual densities to reconstruct the joint density. For example, a credit loss can be viewed as a combination of (1) default, which is a random variable with a value of one for default and zero otherwise, and (2) the exposure, which is a random variable representing the amount at risk, for instance the positive market value of a swap. If the two variables are independent, we can construct the distribution of the credit loss easily. In the case of the two dice, the events are indeed independent. As a result, the probability of a joint event is simply the product of probabilities. For instance, the probability of throwing two ones is equal to $1/6 \times 1/6 = 1/36$.

It is also useful to characterize the distribution of x_1 abstracting from x_2. By integrating over all values of x_2, we obtain the **marginal density**

$$f_1(x_1) = \int_{-\infty}^{\infty} f_{12}(x_1, u_2) du_2 \tag{2.18}$$

and similarly for x_2. We can then define the **conditional density** as

$$f_{1 \cdot 2}(x_1 \mid x_2) = \frac{f_{12}(x_1, x_2)}{f_2(x_2)} \tag{2.19}$$

Here, we keep x_2 fixed and divide the joint density by the marginal probability of x_2. This normalization is necessary to ensure that the conditional density is a proper density function that integrates to one. This relationship is also known as **Bayes' rule.**

2.2.2 Copulas

When the two variables are independent, the joint density is simply the product of the marginal densities. It is rarely the case, however, that financial variables are independent. Dependencies can be modeled by a function called the **copula**, which links, or attaches, marginal distributions into a joint distribution. Formally, the copula is a function of the marginal distributions $F(x)$, plus some parameters, θ, that are specific to this function (and not to the marginals). In the bivariate case, it has two arguments

$$c_{12}[F_1(x_1), F_2(x_2); \theta] \tag{2.20}$$

The link between the joint and marginal distribution is made explicit by *Sklar's theorem*, which states that, for any joint density, there exists a copula that links the marginal densities

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2); \theta] \tag{2.21}$$

With independence, the copula function is a constant always equal to one.

Thus the copula contains all the information on the nature of the dependence between the random variables but gives no information on the marginal distributions. Complex dependencies can be modeled with different copulas. Copulas are now used extensively for modeling financial instruments such as **collateralized debt obligations** (CDOs). As we shall see in a later chapter, CDOs involve movements in many random variables, which are the default events for the companies issuing the debt.

2.2.3 Covariances and Correlations

When dealing with two random variables, the comovement can be described by the **covariance**

$$\text{Cov}(X_1, X_2) = \sigma_{12} = \int_1 \int_2 [x_1 - E(X_1)][x_2 - E(X_2)]\, f_{12}(x_1, x_2) dx_1 dx_2 \tag{2.22}$$

It is often useful to scale the covariance into a unitless number, called the **correlation coefficient**, obtained as

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2} \tag{2.23}$$

The correlation coefficient is a measure of linear dependence. One can show that the correlation coefficient always lies in the $[-1, +1]$ interval. A correlation of one means that the two variables always move in the same direction. A correlation of minus one means that the two variables always move in opposite direction.

If the variables are independent, the joint density separates out and this becomes

$$\mathrm{Cov}(X_1, X_2) = \left\{\int_1 [x_1 - E(X_1)] f_1(x_1) dx_1\right\} \left\{\int_2 [x_2 - E(X_2)] f_2(x_2) dx_2\right\} = 0$$

by Equation (2.6), since the average deviation from the mean is zero. In this case, the two variables are said to be **uncorrelated.** Hence independence implies zero correlation (the reverse is not true, however).

Example: Multivariate Functions

Consider two variables, such as the exchange rates for the Canadian dollar and the euro. Table 2.3a describes the joint density function $f_{12}(x_1, x_2)$, assuming two payoffs only for each variable. Note first that the density indeed sums to $0.30 + 0.20 + 0.15 + 0.35 = 1.00$.

TABLE 2.3a Joint Density Function

x_1 x_2	-5	$+5$
-10	0.30	0.15
$+10$	0.20	0.35

From this, we can compute the marginal density for each variable, along with its mean and standard deviation. For instance, the marginal probability of $x_1 = -5$ is given by $f_1(x_1) = f_{12}(x_1, x_2 = -10) + f_{12}(x_1, x_2 = +10) = 0.30 + 0.20 = 0.50$. The marginal probability of $x_1 = +5$ must be 0.50 as well. Table 2.3b shows that the means and standard deviations are, respectively, $\bar{x}_1 = 0.0, \sigma_1 = 5.0$, and $\bar{x}_2 = 1.0, \sigma_2 = 9.95$.

Finally, Table 2.3c details the computation of the covariance, which gives $\mathrm{Cov} = 15.00$. Dividing by the product of the standard deviations, we get $\rho = \mathrm{Cov}/(\sigma_1\sigma_2) = 15.00/(5.00 \times 9.95) = 0.30$. The positive correlation indicates that when one variable goes up, the other is more likely to go up than down.

TABLE 2.3b Marginal Density Functions

Variable 1				Variable 2			
x_1	Prob. $f_1(x_1)$	Mean $x_1 f_1(x_1)$	Variance $(x_1 - \bar{x}_1)^2 f_1(x_1)$	x_2	Prob. $f_2(x_2)$	Mean $x_2 f_2(x_2)$	Variance $(x_2 - \bar{x}_2)^2 f_2(x_2)$
-5	0.50	-2.5	12.5	-10	0.45	-4.5	54.45
$+5$	0.50	$+2.5$	12.5	$+10$	0.55	$+5.5$	44.55
Sum	1.00	0.0	25.0	Sum	1.00	1.0	99.0
		$\bar{x}_1 = 0.0$	$\sigma_1 = 5.0$			$\bar{x}_2 = 1.0$	$\sigma_2 = 9.95$

TABLE 2.3c Covariance and Correlation

	$(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) f_{12}(x_1, x_2)$	
	$x_1 = -5$	$x_1 = +5$
$x_2 = -10$	$(-5 - 0)(-10 - 1)0.30 = 16.50$	$(+5 - 0)(-10 - 1)0.15 = -8.25$
$x_2 = +10$	$(-5 - 0)(+10 - 1)0.20 = -9.00$	$(+5 - 0)(+10 - 1)0.35 = 15.75$
Sum	Cov=15.00	

EXAMPLE 2.1: FRM EXAM 2000—QUESTION 81

Which one of the following statements about the correlation coefficient is *false*?

a. It always ranges from -1 to $+1$.
b. A correlation coefficient of zero means that two random variables are independent.
c. It is a measure of linear relationship between two random variables.
d. It can be calculated by scaling the covariance between two random variables.

EXAMPLE 2.2: FRM EXAM 2007—QUESTION 93

The joint probability distribution of random variables X and Y is given by $f(x, y) = k \times x \times y$ for $x = 1, 2, 3$, $y = 1, 2, 3$, and k is a positive constant. What is the probability that $X + Y$ will exceed 5?

a. 1/9
b. 1/4
c. 1/36
d. Cannot be determined

2.3 FUNCTIONS OF RANDOM VARIABLES

Risk management is about uncovering the distribution of portfolio values. Consider a security that depends on a unique source of risk, such as a bond. The risk manager could model the change in the bond price as a random variable directly. The problem with this choice is that the distribution of the bond price is not stationary, because the price converges to the face value at expiration.

Instead, the practice is to model the change in yields as a random variable because its distribution is better behaved. The next step is to use the relationship between the bond price and the yield to uncover the distribution of the bond price.

This illustrates a general principle of risk management, which is to model the risk factor first, then to derive the distribution of the instrument from information about the function that links the instrument value to the risk factor. This may not be easy to do, unfortunately, if the relationship is highly nonlinear. In what follows, we first focus on the mean and variance of simple transformations of random variables.

2.3.1 Linear Transformation of Random Variables

Consider a transformation that multiplies the original random variable by a constant and add a fixed amount, $Y = a + bX$. The expectation of Y is

$$E(a + bX) = a + bE(X) \tag{2.24}$$

and its variance is

$$V(a + bX) = b^2 V(X) \tag{2.25}$$

Note that adding a constant never affects the variance since the computation involves the *difference* between the variable and its mean. The standard deviation is

$$\mathrm{SD}(a + bX) = b\mathrm{SD}(X) \tag{2.26}$$

Example: Currency Position Plus Cash

A dollar-based investor has a portfolio consisting of \$1 million in cash plus a position in 1,000 million Japanese yen. The distribution of the dollar/yen exchange rate X has a mean of $E(X) = 1/100 = 0.01$ and volatility of $\mathrm{SD}(X) = 0.10/100 = 0.001$.

The portfolio value can be written as $Y = a + bX$, with fixed parameters (in millions) $a = \$1$ and $b = Y1,000$. Therefore, the portfolio expected value is $E(Y) = \$1 + Y1,000 \times 1/100 = \$11$ million, and the standard deviation is $\mathrm{SD}(Y) = Y1,000 \times 0.001 = \$1$ million.

2.3.2 Sum of Random Variables

Another useful transformation is the summation of two random variables. A portfolio, for instance, could contain one share of Intel plus one share of Microsoft. The rate of return on each stock behaves as a random variable.

The expectation of the sum $Y = X_1 + X_2$ can be written as

$$E(X_1 + X_2) = E(X_1) + E(X_2) \tag{2.27}$$

and its variance is

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2\mathrm{Cov}(X_1, X_2) \tag{2.28}$$

When the variables are uncorrelated, the variance of the sum reduces to the sum of variances. Otherwise, we have to account for the cross-product term.

KEY CONCEPT

The expectation of a sum is the sum of expectations. The variance of a sum, however, is only the sum of variances if the variables are uncorrelated.

2.3.3 Portfolios of Random Variables

More generally, consider a linear combination of a number of random variables. This could be a portfolio with fixed weights, for which the rate of return is

$$Y = \sum_{i=1}^{N} w_i X_i \tag{2.29}$$

where N is the number of assets, X_i is the rate of return on asset i, and w_i its weight.

To shorten notation, this can be written in matrix notation, replacing a string of numbers by a single vector:

$$Y = w_1 X_1 + w_2 X_2 + \cdots + w_N X_N = \left[w_1 w_2 \ldots w_N\right] \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = w'X \tag{2.30}$$

where w' represents the transposed vector (i.e., horizontal) of weights and X is the vertical vector containing individual asset returns. The appendix for this chapter provides a brief review of matrix multiplication.

The portfolio expected return is now

$$\mathrm{E}(Y) = \mu_p = \sum_{i=1}^{N} w_i \mu_i \tag{2.31}$$

which is a weighted average of the expected returns $\mu_i = E(X_i)$. The variance is

$$V(Y) = \sigma_p^2 = \sum_{i=1}^{N} w_i^2\sigma_i^2 + \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} w_i w_j \sigma_{ij} = \sum_{i=1}^{N} w_i^2\sigma_i^2 + 2\sum_{i=1}^{N}\sum_{j<i}^{N} w_i w_j \sigma_{ij} \tag{2.32}$$

Using matrix notation, the variance can be written as

$$\sigma_p^2 = [w_1 \dots w_N] \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1N} \\ \vdots & & & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \dots & \sigma_N \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix}$$

Defining Σ as the covariance matrix, the variance of the portfolio rate of return can be written more compactly as

$$\sigma_p^2 = w'\Sigma w \tag{2.33}$$

This is a useful expression to describe the risk of the total portfolio.

Example: Computing the Risk of a Portfolio

Consider a portfolio invested in Canadian dollars and euros. The joint density function is given by Table 2.3a. Here, x_1 describes the payoff on the Canadian dollar, with $\mu_1 = 0.00$, $\sigma_1 = 5.00$, and $\sigma_1^2 = 25$. For the euro, $\mu_2 = 1.00$, $\sigma_2 = 9.95$, and $\sigma_2^2 = 99$. The covariance was computed as $\sigma_{12} = 15.00$, with the correlation $\rho = 0.30$. If we have 60% invested in Canadian dollar and 40% in euros, what is the portfolio volatility?

Following Equation (2.33), we write

$$\sigma_p^2 = [0.60\ 0.40] \begin{bmatrix} 25 & 15 \\ 15 & 99 \end{bmatrix} \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix} = [0.60\ 0.40] \begin{bmatrix} 25 \times 0.60 + 15 \times 0.40 \\ 15 \times 0.60 + 99 \times 0.40 \end{bmatrix}$$

$$\sigma_p^2 = [0.60\ 0.40] \begin{bmatrix} 21.00 \\ 48.60 \end{bmatrix} = 0.60 \times 21.00 + 0.40 \times 48.60 = 32.04$$

Therefore, the portfolio volatility is $\sigma_p = \sqrt{32.04} = 5.66$. Note that this is hardly higher than the volatility of the Canadian dollar alone, even though the risk of the euro is much higher. The portfolio risk has been kept low due to a diversification effect, or low correlation between the two assets.

2.3.4 Product of Random Variables

Some risks result from the product of two random variables. A credit loss, for instance, arises from the product of the occurrence of default and the loss given default.

Using Equation (2.22), the expectation of the product $Y = X_1 X_2$ can be written as

$$E(X_1\ X_2) = E(X_1)E(X_2) + \mathrm{Cov}(X_1, X_2) \tag{2.34}$$

When the variables are independent, this reduces to the product of the means.

The variance is more complex to evaluate. With independence, it reduces to:

$$V(X_1\ X_2) = E(X_1)^2 V(X_2) + V(X_1)E(X_2)^2 + V(X_1)V(X_2) \tag{2.35}$$

2.3.5 Distributions of Transformations of RVs

The preceding results focus on the mean and variance of simple transformations only. They do not fully describe the distribution of the transformed variable $Y = g(X)$. This, unfortunately, is usually complicated for all but the simplest transformations $g(\cdot)$ and densities $f(X)$.

Even if there is no closed-form solution for the density, we can describe the cumulative distribution function of Y when $g(X)$ is a one-to-one transformation from X into Y. This implies that the function can be inverted, or that for a given y, we can find x such that $x = g^{-1}(y)$. We can then write

$$P[Y \le y] = P[g(X) \le y] = P[X \le g^{-1}(y)] = F_X(g^{-1}(y)) \tag{2.36}$$

where $F(\cdot)$ is the cumulative distribution function of X. Here, we assumed the relationship is positive. Otherwise, the right-hand term is changed to $1 - F_X(g^{-1}(y))$.

This allows us to derive the quantile of, say, the bond price from information about the probability distribution of the yield. Suppose we consider a zero-coupon bond, for which the market value V is

$$V = \frac{100}{(1+r)^T} \tag{2.37}$$

where r is the yield. This equation describes V as a function of r, or $Y = g(X)$. Using $r = 6\%$ and $T = 30$ years, the current price is $V = \$17.41$. The inverse function $X = g^{-1}(Y)$ is

$$r = (100/V)^{1/T} - 1 \tag{2.38}$$

We wish to estimate the probability that the bond price could fall below a cutoff price $V = \$15$. We invert the price-yield function and compute the associated yield level, $g^{-1}(y) = (100/\$15)^{1/30} - 1 = 6.528\%$. Lower prices are associated with higher yield levels. Using Equation (2.36), the probability is given by

$$P[V \le \$15] = P[r \ge 6.528\%]$$

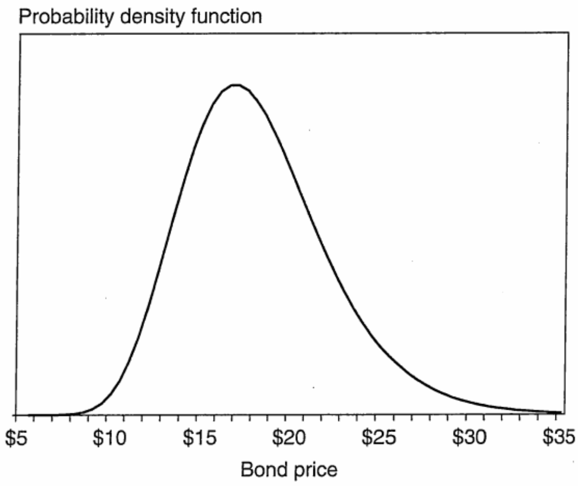

FIGURE 2.5 Density Function for the Bond Price

Assuming the yield change is normal with volatility 0.8%, this gives a probability of 25.5 percent.[1] Even though we do not know the density of the bond price, this method allows us to trace out its cumulative distribution by changing the cutoff price of \$15. Taking the derivative, we can recover the density function of the bond price. Figure 2.5 shows that this p.d.f. is skewed to the right.

On the extreme right, if the yield falls to zero, the bond price will go to \$100. On the extreme left, if the yield goes to infinity, the bond price will fall to, but not go below, zero. Relative to the current value of \$17.41, there is a greater likelihood of large movements up than down.

This method, unfortunately, cannot be easily extended. For general density functions and transformations, risk managers turn to numerical methods, especially when the number of random variables is large. This is why credit risk models, for instance, all describe the distribution of credit losses through simulations.

EXAMPLE 2.3: FRM EXAM 2007—QUESTION 127

Suppose that A and B are random variables, each follows a standard normal distribution, and the covariance between A and B is 0.35. What is the variance of $(3A + 2B)$?

a. 14.47
b. 17.20
c. 9.20
d. 15.10

[1]We shall see later that this is obtained from the standard normal variable $z = (6.528 - 6.000)/0.80 = 0.660$. Using standard normal tables, or the NORMSDIST(−0.660) Excel function, this gives 25.5%.

EXAMPLE 2.4: FRM EXAM 2002—QUESTION 70

Given that x and y are random variables, and a, b, c and d are constant, which one of the following definitions is *wrong*.

a. $E(ax + by + c) = aE(x) + bE(y) + c$, if x and y are correlated.
b. $V(ax + by + c) = V(ax + by) + c$, if x and y are correlated.
c. $Cov(ax + by, cx + dy) = acV(x) + bdV(y) + (ad + bc)Cov(x, y)$, if x and y are correlated.
d. $V(x - y) = V(x + y) = V(x) + V(y)$, if x and y are uncorrelated.

2.4 IMPORTANT DISTRIBUTION FUNCTIONS

2.4.1 Uniform Distribution

The simplest continuous distribution function is the **uniform distribution**. This is defined over a range of values for x, $a \le x \le b$. The density function is

$$f(x) = \frac{1}{(b-a)}, \quad a \le x \le b \tag{2.39}$$

which is constant and indeed integrates to unity. This distribution puts the same weight on each observation within the allowable range, as shown in Figure 2.6. We denote this distribution as $U(a, b)$.

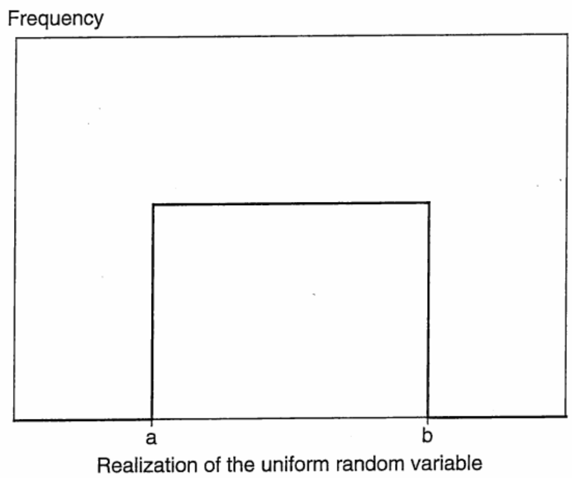

FIGURE 2.6 Uniform Density Function

Its mean and variance are given by

$$E(X) = \frac{a+b}{2} \tag{2.40}$$

$$V(X) = \frac{(b-a)^2}{12} \tag{2.41}$$

The uniform distribution $U(0, 1)$ is widely used as a starting distribution for generating random variables from any distribution $F(Y)$ in simulations. We need to have analytical formulas for the p.d.f. $f(Y)$ and its cumulative distribution $F(Y)$. As any cumulative distribution function ranges from zero to unity, we first draw X from $U(0, 1)$ and then compute $y = F^{-1}(x)$. The random variable Y will then have the desired distribution $f(Y)$.

EXAMPLE 2.5: FRM EXAM 2002—QUESTION 119

The random variable X with density function $f(x) = 1/(b-a)$ for $a < x < b$, and 0 otherwise, is said to have a uniform distribution over (a, b). Calculate its mean.

a. $(a+b)/2$
b. $a - b/2$
c. $a + b/4$
d. $a - b/4$

2.4.2 Normal Distribution

Perhaps the most important continuous distribution is the **normal distribution,** which represents adequately many random processes. This has a bell-like shape with more weight in the center and tails tapering off to zero. The daily rate of return in a stock price, for instance, has a distribution similar to the normal p.d.f.

The normal distribution can be characterized by its first two moments only, the mean μ and variance σ^2. The first parameter represents the location; the second, the dispersion. The normal density function has the following expression:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \tag{2.42}$$

Its mean is $E[X] = \mu$ and variance $V[X] = \sigma^2$. We denote this distribution as $N(\mu, \sigma^2)$. Because the function can be fully specified by these two parameters, it is called a **parametric function.**

Instead of having to deal with different parameters, it is often more convenient to use a **standard normal variable** as ϵ, which has been standardized, or

TABLE 2.4 Lower Quantiles of the Standardized Normal Distribution

	Confidence Level (percent)								
	99.99	99.9	99	97.72	97.5	95	90	84.13	50
$c(-\alpha)$	−3.715	−3.090	−2.326	−2.000	−1.960	−1.645	−1.282	−1.000	−0.000

normalized, so that $E(\epsilon) = 0$, $V(\epsilon) = \sigma(\epsilon) = 1$. Figure 2.7 plots the **standard normal density.**

First, note that the function is symmetrical around the mean. Its mean of zero is the same as its **mode** (which is also the most likely, or highest, point on this curve) and **median** (which is such that the area to the left is a 50 percent probability). The skewness of a normal distribution is 0, which indicates that it is symmetrical around the mean. The kurtosis of a normal distribution is 3. Distributions with fatter tails have a greater kurtosis coefficient.

About 95% of the distribution is contained between values of $\epsilon_1 = -2$ and $\epsilon_2 = +2$, and 68% of the distribution falls between values of $\epsilon_1 = -1$ and $\epsilon_2 = +1$. Table 2.4 gives the values that correspond to right-tail probabilities, such that

$$\int_{-\alpha}^{\infty} f(\epsilon)d\epsilon = c \tag{2.43}$$

For instance, the value of −1.645 is the quantile that corresponds to a 95% probability.[2]

This distribution plays a central role in finance because it represents adequately the behavior of many financial variables. It enters, for instance, the Black–Scholes

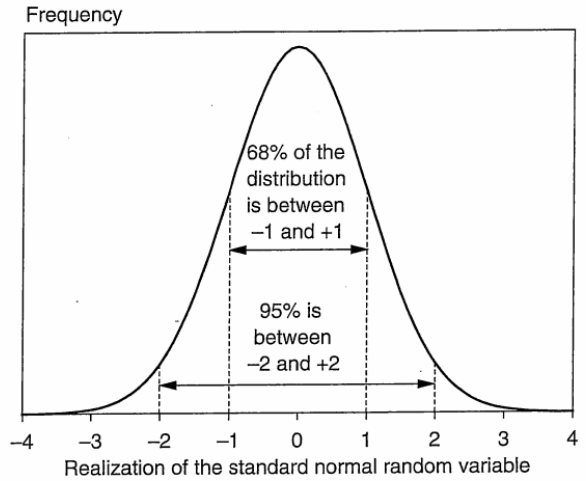

FIGURE 2.7 Normal Density Function

[2] More generally, the cumulative distribution can be found from the Excel function NORMDIST(·). For example, we can verify that NORMSDIST(−1.645) yields 0.04999, or a 5% left-tail probability.

option pricing formula where the function $N(\cdot)$ represents the cumulative standardized normal distribution function.

The distribution of any normal variable can then be recovered from that of the standard normal, by defining

$$X = \mu + \epsilon\sigma \tag{2.44}$$

Using Equations (2.24) and (2.25), we can show that X has indeed the desired moments, as $E(X) = \mu + E(\epsilon)\sigma = \mu$ and $V(X) = V(\epsilon)\sigma^2 = \sigma^2$.

Define, for instance, the random variable as the change in the dollar value of a portfolio. The expected value is $E(X) = \mu$. To find the quantile of X at the specified confidence level c, we replace ϵ by $-\alpha$ in Equation (2.44). This gives $Q(X, c) = \mu - \alpha\sigma$. Using Equation (2.9), we can compute VAR as

$$\text{VAR} = E(X) - Q(X, c) = \mu - (\mu - \alpha\sigma) = \alpha\sigma \tag{2.45}$$

For example, a portfolio with a standard deviation of \$10 million would have a VAR, or potential downside loss, of \$16.45 million at the 95% confidence level.

KEY CONCEPT

With normal distributions, the VAR of a portfolio is obtained from the product of the portfolio standard deviation and a standard normal deviate factor that reflects the confidence level, for instance 1.645 at the 95% level.

An important property of the normal distribution is that it is one of the few distributions that is *stable* under addition. In other words, a linear combination of jointly normally distributed random variables has a normal distribution.[3] This is extremely useful because we only need to know the mean and variance of the portfolio to reconstruct its whole distribution.

KEY CONCEPT

A linear combination of jointly normal variables has a normal distribution.

[3] Strictly speaking, this is only true under either of the following conditions: (1) the univariate variables are independently distributed, or (2) the variables are multivariate normally distributed (this invariance property also holds for jointly elliptically distributed variables).

When we have N random variables, the joint normal density can be written as a function of the vector x, of the means μ, and the covariance matrix Σ:

$$f(x_1, \cdots, x_N) = \frac{1}{(2\pi)^{N/2} \mid \Sigma \mid^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma(x-\mu)\right] \tag{2.46}$$

Using the concept of copulas, this can be separated into N different marginal normal densities and a joint normal copula. For two random variables, Equation (2.21) showed

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2); \theta]$$

Here, both f_1 and f_2 are normal marginals. They have parameters μ_1 and σ_1, and μ_2 and σ_2. In addition, c_{12} is the normal copula. Note that its sole parameter is the correlation coefficient ρ_{12}. This additional information is required to construct the covariance matrix Σ and defines the strength of the dependency between the two variables.

EXAMPLE 2.6: FRM EXAM 2005—QUESTION 62

Let Z be a standard normal random variable. An event X is defined to happen if either Z takes a value between -0.5 and $+0.5$ or Z takes any value greater than 1.5. What is the probability of event X happening if $N(0.5) = 0.6915$ and $N(-1.5) = 0.0668$, where $N(.)$ is the cumulative distribution function of a standard normal variable?

a. 0.2583
b. 0.3753
c. 0.4498
d. 0.7583

EXAMPLE 2.7: FRM EXAM 2003—QUESTION 21

Which of the following statements about the normal distribution is *not* accurate?

a. Kurtosis equals 3.
b. Skewness equals 1.
c. The entire distribution can be characterized by two moments, mean and variance.
d. The normal density function has the following expression: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2]$

EXAMPLE 2.8: FRM EXAM 2006—QUESTION 11

Which type of distribution produces the lowest probability for a variable to exceed a specified extreme value which is greater than the mean, assuming the distributions all have the same mean and variance?

a. A leptokurtic distribution with a kurtosis of 4
b. A leptokurtic distribution with a kurtosis of 8
c. A normal distribution
d. A platykurtic distribution

2.4.3 Lognormal Distribution

The normal distribution is a good approximation for many financial variables, such as the rate of return on a stock, $r = (P_1 - P_0)/P_0$, where P_0 and P_1 are the stock prices at time 0 and 1.

Strictly speaking, this is inconsistent with reality since a normal variable has infinite tails on both sides. In theory, r could end up below -1, which implies $P_1 < 0$. In reality, due to the limited liability of corporations, stock prices cannot turn negative. In many situations, however, this is an excellent approximation. For instance, with short horizons or small price moves, the probability of having a negative price is so small that it is negligible. If this is not the case, we need to resort to other distributions that prevent prices from going negative. One such distribution is the lognormal.

A random variable X is said to have a **lognormal distribution** if its logarithm $Y = \ln(X)$ is normally distributed. Define here $X = (P_1/P_0)$. Because the argument X in the logarithm function must be positive, the price P_1 can never go below zero.

The lognormal density function has the following expression

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2\right], \quad x > 0 \tag{2.47}$$

Note that this is more complex than simply plugging $\ln(x)$ in Equation (2.42), because x also appears in the denominator. Its mean is

$$E[X] = \exp\left[\mu + \frac{1}{2}\sigma^2\right] \tag{2.48}$$

and variance $V[X] = \exp[2\mu + 2\sigma^2] - \exp[2\mu + \sigma^2]$. The parameters were chosen to correspond to those of the normal variable, $E[Y] = E[\ln(X)] = \mu$ and $V[Y] = V[\ln(X)] = \sigma^2$.

Conversely, if we set $E[X] = \exp[r]$, the mean of the associated normal variable is $E[Y] = E[\ln(X)] = (r - \sigma^2/2)$. We will see later that this adjustment is also

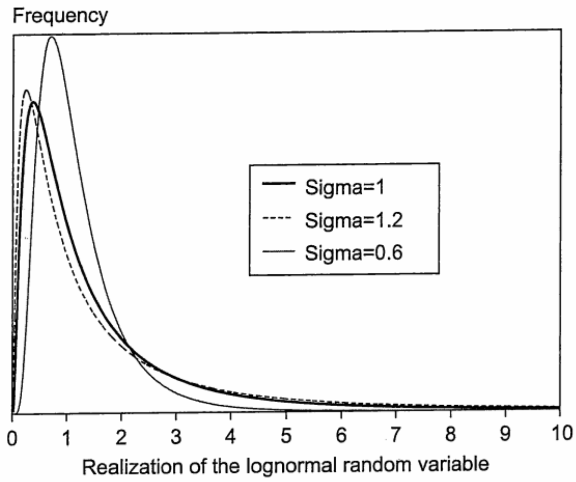

FIGURE 2.8 Lognormal Density Function

used in the Black–Scholes option valuation model, where the formula involves a trend in $(r - \sigma^2/2)$ for the log-price ratio.

Figure 2.8 depicts the lognormal density function with $\mu = 0$, and various values $\sigma = 1.0, 1.2, 0.6$. Note that the distribution is skewed to the right. The tail increases for greater values of σ. This explains why as the variance increases, the mean is pulled up in Equation (2.48).

We also note that the distribution of the bond price in our previous example, Equation (2.37), resembles a lognormal distribution. Using continuous compounding instead of annual compounding, the price function is

$$V = 100 \exp(-rT) \tag{2.49}$$

which implies $\ln(V/100) = -rT$. Thus if r is normally distributed, V has a lognormal distribution.

EXAMPLE 2.9: FRM EXAM 1999—QUESTION 5

Which of the following statements best characterizes the relationship between the normal and lognormal distributions?

a. The lognormal distribution is the logarithm of the normal distribution.
b. If the natural log of the random variable X is lognormally distributed, then X is normally distributed.
c. If X is lognormally distributed, then the natural log of X is normally distributed.
d. The two distributions have nothing to do with one another.

EXAMPLE 2.10: FRM EXAM 2007—QUESTION 21

The skew of a lognormal distribution is always

a. Positive
b. Negative
c. 0
d. 3

EXAMPLE 2.11: FRM EXAM 2002—QUESTION 125

Consider a stock with an initial price of \$100. Its price one year from now is given by $S = 100 \times \exp(r)$, where the rate of return r is normally distributed with a mean of 0.1 and a standard deviation of 0.2. With 95% confidence, after rounding, S will be between

a. \$67.57 and \$147.99
b. \$70.80 and \$149.20
c. \$74.68 and \$163.56
d. \$102.18 and \$119.53

EXAMPLE 2.12: FRM EXAM 2000—QUESTION 128

For a lognormal variable X, we know that $\ln(X)$ has a normal distribution with a mean of zero and a standard deviation of 0.5. What are the expected value and and the variance of X?

a. 1.025 and 0.187
b. 1.126 and 0.217
c. 1.133 and 0.365
d. 1.203 and 0.399

2.4.4 Student's *t* Distribution

Another important distribution is the **Student's *t* distribution**. This arises in hypothesis testing, because it describes the distribution of the ratio of the estimated coefficient to its standard error.

This distribution is characterized by a parameter k known as the **degrees of freedom.** Its density is

$$f(x) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+x^2/k)^{(k+1)/2}} \tag{2.50}$$

where Γ is the gamma function, defined as $\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx$. As k increases, this function converges to the normal p.d.f.

The distribution is symmetrical with mean zero and variance

$$V[X] = \frac{k}{k-2} \tag{2.51}$$

provided $k > 2$. Its kurtosis is

$$\delta = 3 + \frac{6}{k-4} \tag{2.52}$$

provided $k > 4$. Its has fatter tails than the normal which often provides a better representation of typical financial variables. Typical estimated values of k are around four to six for stock returns. Figure 2.9 displays the density for $k = 4$ and $k = 50$. The latter is close to the normal. With $k = 4$, however, the p.d.f. has fatter tails. As was done for the normal density, we can also use the Student's t to compute VAR as a function of the volatility

$$\text{VAR} = \alpha_k \sigma \tag{2.53}$$

where the multiplier now depends on the degrees of freedom k.

As for the multivariate normal distribution, the joint Student distribution can be separated into two components. The marginals have the Student's distribution

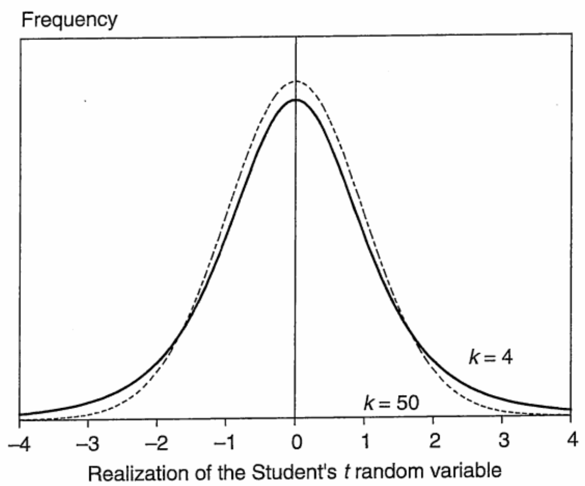

FIGURE 2.9 Student's t Density Function

described in Equation (2.50). In addition, the copula has a specific shape, which is the Student's copula. This copula allows for stronger dependencies in the tails than the normal copula. Marginals and copulas of different types can be used, as best fits the data. For example, one could use normal marginals and a Student's copula. This creates substantial flexibility in the statistical modeling of random variables.

Another distribution derived from the normal is the **chi-square distribution,** which can be viewed as the sum of independent squared standard normal variables

$$x = \sum_{j=1}^{k} z_j^2 \tag{2.54}$$

where k is also called the degrees of freedom. Its mean is $E[X] = k$ and variance $V[X] = 2k$. For k sufficiently large, $\chi^2(k)$ converges to a normal distribution $N(k, 2k)$. This distribution describes the sample variance.

Finally, another associated distribution is the F **distribution,** which can be viewed as the ratio of independent chi-square variables divided by their degrees of freedom

$$F(a, b) = \frac{\chi^2(a)/a}{\chi^2(b)/b} \tag{2.55}$$

This distribution appears in joint tests of regression coefficients.

EXAMPLE 2.13: FRM EXAM 2003—QUESTION 18

Which of the following statements is the most accurate about the relationship between a normal distribution and a Student's t-distribution that have the same mean and standard deviation?

a. They have the same skewness and the same kurtosis.
b. The Student's t-distribution has larger skewness and larger kurtosis.
c. The kurtosis of a Student's t-distribution converges to that of the normal distribution as the number of degrees of freedom increases.
d. The normal distribution is a good approximation for the Student's t-distribution when the number of degrees of freedom is small.

2.4.5 Binomial Distribution

Consider now a random variable that can take discrete values between zero and n. This could be, for instance, the number of times VAR is exceeded over the last year, also called the number of **exceptions.** Thus, the binomial distribution plays an important role for the backtesting of VAR models.

A binomial variable can be viewed as the result of n independent **Bernoulli trials**, where each trial results in an outcome of $y = 0$ or $y = 1$. This applies, for example, to credit risk. In case of default, we have $y = 1$, otherwise $y = 0$. Each Bernoulli variable has expected value of $E[Y] = p$ and variance $V[Y] = p(1 - p)$.

A random variable is defined to have a **binomial distribution** if the discrete density function is given by

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n \tag{2.56}$$

where $\binom{n}{x}$ is the number of combinations of n things taken x at a time, or

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{2.57}$$

and the parameter p is between zero and one. This distribution also represents the total number of successes in n repeated experiments where each success has a probability of p.

The binomial variable has mean and variance

$$E[X] = pn \tag{2.58}$$

$$V[X] = p(1-p)n \tag{2.59}$$

It is described in Figure 2.10 in the case where $p = 0.25$ and $n = 10$. The probability of observing $X = 0, 1, 2 \ldots$ is 5.6%, 18.8%, 28.1% and so on.

For instance, we want to know what is the probability of observing $x = 0$ exceptions out of a sample of $n = 250$ observations when the true probability is 1%. We should expect to observe 2.5 exceptions on average across many such samples. There will be, however, some samples with no exceptions at all. This

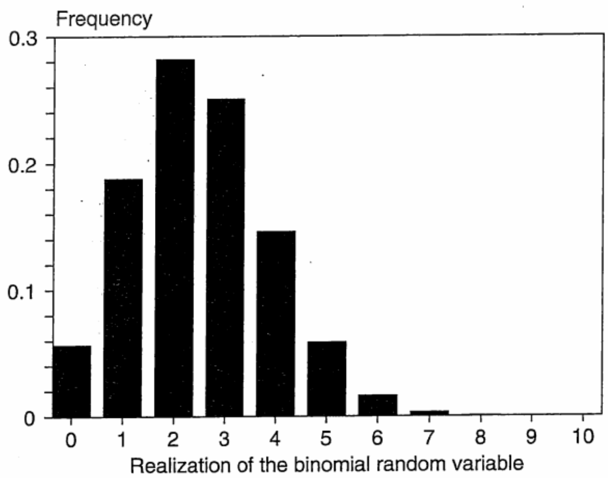

FIGURE 2.10 Binomial Density Function with $p = 0.25, n = 10$

probability is

$$f(X=0) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} = \frac{250!}{1 \times 250!}0.01^0 0.99^{250} = 0.081$$

So, we would expect to observe 8.1% of samples with zero exceptions, under the null hypothesis. We can repeat this calculation with different values for x. For example, the probability of observing 8 exceptions is $f(X=8) = 0.02\%$ only. We can use this information to test the null hypothesis. Because this probability is so low, observing 8 exceptions would make us question whether the true probability is 1%.

EXAMPLE 2.14: FRM EXAM 2006—QUESTION 84

On a multiple-choice exam with four choices for each of six questions, what is the probability that a student gets fewer than two questions correct simply by guessing?

a. 0.46%
b. 23.73%
c. 35.60%
d. 53.39%

2.4.6 Poisson Distribution

The Poisson distribution is a discrete distribution, which typically is used to describe the number of events occurring over a fixed period of time, assuming events are independent of each other. It is defined as

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots \tag{2.60}$$

where λ is a positive number representing the average arrival rate during the period. This distribution, for example, is widely used to represent the frequency, or number of occurrences, of operational losses over a year.

The parameter λ represents the expected value of X and also its variance

$$E[X] = \lambda \tag{2.61}$$

$$V[X] = \lambda \tag{2.62}$$

The Poisson distribution is the limiting case of the binomial distribution as n goes to infinity and p goes to zero, while $np = \lambda$ remains fixed. In addition, when λ

is large the Poisson distribution is well approximated by the normal distribution with mean and variance of λ, through the central limit theorem.

If the number of arrivals follows a Poisson distribution, then the time period between arrivals follows an **exponential distribution** with mean $1/\lambda$. The latter has density taking the form $f(x) = \lambda e^{-\lambda x}$, for $x \geq 0$. For example, if we expect $\lambda = 12$ losses per year, the average time interval between losses should be 1 year divided by 12, or one month.

EXAMPLE 2.15: FRM EXAM 2004—QUESTION 60

When can you use the normal distribution to approximate the Poisson distribution, assuming you have n independent trials each with a probability of success of p

a. When the mean of the Poisson distribution is very small
b. When the variance of the Poisson distribution is very small
c. When the number of observations is very large and the success rate is close to 1
d. When the number of observations is very large and the success rate is close to 0

2.5 LIMIT DISTRIBUTIONS

2.5.1 Distribution of Averages

The normal distribution is extremely important because of the **central limit theorem** (CLT), which states that the mean of n independent and identically distributed variables converges to a normal distribution as the number of observations n increases. This very powerful result is valid for any underlying distribution, as long as the realizations are independent. For instance, the distribution of total credit losses converges to a normal distribution as the number of loans increases to a large value, assuming defaults are always independent of each other.

Define $\bar{X}$ as the mean $\frac{1}{n}\sum_{i=1}^{n} X_i$, where each variable has mean μ and standard deviation σ. We have

$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right) \tag{2.63}$$

Standardizing the variable, we can write

$$\frac{\bar{X} - \mu}{(\sigma/\sqrt{n})} \rightarrow N(0, 1) \tag{2.64}$$

Thus, the normal distribution is the limiting distribution of the average, which explain why it has such a prominent place in statistics.

As an example, consider the binomial variable, which is the sum of independent Bernoulli trials. When n is large, we can use the CLT and approximate the binomial distribution by the normal distribution. Using Equation (2.64) for the sum, we have

$$z = \frac{x - pn}{\sqrt{p(1-p)n}} \rightarrow N(0, 1) \tag{2.65}$$

which is much easier to evaluate than the binomial distribution.

Consider for example the issue of whether the number of exceptions x we observe is compatible with a 99% VAR. For our example, the mean and variance of x are $E[X] = 0.01 \times 250 = 2.5$ and $V[X] = 0.01(1 - 0.01) \times 250 = 2.475$. We observe $x = 8$, which gives $z = (8 - 2.5)/\sqrt{2.475} = 3.50$. We can now compare this number to the standard normal distribution. Say for instance that we decide to reject the hypothesis that VAR is correct if the statistic falls outside a 95% two-tailed confidence band.[4] This interval is $(-1.96, +1.96)$ for the standardized normal distribution. Here, the value of 3.50 is much higher than the cutoff point of $+1.96$. As a result, we would reject the null hypothesis that the true probability of observing an exception is 1% only. In other words, there are simply too many exceptions to be explained by bad luck. It is more likely that the VAR model underestimates risk.

2.5.2 Distribution of Tails

The CLT deals with the mean, or center of the distribution. For risk management purposes, it is also useful to examine the tails of the distribution.

Another powerful theorem is given by **extreme value theory** (EVT). The EVT theorem says that the limit distribution for values x beyond a cutoff point u belongs to the following family

$$\begin{aligned} F(y) &= 1 - (1 + \xi y)^{-1/\xi}, \quad && \xi \neq 0 \\ F(y) &= 1 - exp(-y), && \xi = 0 \end{aligned} \tag{2.66}$$

where $y = (x - u)/\beta$. To simplify, we define the loss x as a positive number so that y is also positive. The distribution is characterized by $\beta > 0$, a *scale* parameter, and by ξ, a *shape* parameter that determines the speed at which the tail disappears.

This distribution is called the **Generalized Pareto Distribution**, because it subsumes other distributions as special cases. For instance, the normal distribution corresponds to $\xi = 0$, in which case the tails disappear at an exponential speed. Typical financial data have $\xi > 0$, which implies *fat tails*. This class of distribution

[4] Note that the choice of this confidence level has nothing to do with the VAR confidence level. Here, the 95% level represents the rate at which the decision rule will commit the error of falsely rejecting a correct model.

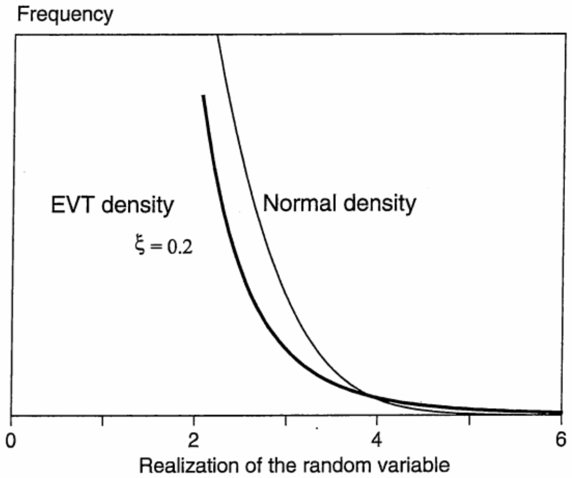

FIGURE 2.11 EVT and Normal Densities

includes the Gumbel, Fréchet, and Weibull families, as $\xi \to 0$, $\xi > 0$, and $\xi < 0$, respectively.

Figure 2.11 illustrates the shape of the density function for U.S. stock market data. The normal density falls off fairly quickly. With $\xi = 0.2$, the EVT density has a fatter tail than the normal density, implying a higher probability of experiencing large losses. This is an important observation for risk management purposes. Note that the EVT density is only defined for the tail, i.e., when the loss x exceeds an arbitrary cutoff point, which is taken as 2 in this case.

EXAMPLE 2.16: FRM EXAM 2007—QUESTION 110

Which of the following statements regarding extreme value theory (EVT) is *incorrect*?

a. In contrast to conventional approaches for estimating VAR, EVT only considers the tail behavior of the distribution.

b. Conventional approaches for estimating VAR that assume that the distribution of returns follows a unique distribution for the entire range of values may fail to properly account for the fat tails of the distribution of returns.

c. EVT attempts to find the optimal point beyond which all values belong to the tail and then models the distribution of the tail separately.

d. By smoothing the tail of the distribution, EVT effectively ignores extreme events and losses that can generally be labeled outliers.

2.6 IMPORTANT FORMULAS

Probability density function: $f(x) = \text{Prob}(X = x)$
(Cumulative) distribution function: $F(x) = \int_{-\infty}^{x} f(u)du$
Mean: $E(X) = \mu = \int x f(x)dx$
Variance: $V(X) = \sigma^2 = \int [x - \mu]^2 f(x)dx$
Skewness: $\gamma = \left(\int [x - \mu]^3 f(x)dx\right) / \sigma^3$
Kurtosis: $\delta = \left(\int [x - \mu]^4 f(x)dx\right) / \sigma^4$
Quantile, VAR: $\text{VAR} = E(X) - Q(X, c) = \alpha\sigma$
Independent joint densities: $f_{12}(x_1 x_2) = f_1(x_1) \times f_2(x_2)$
Marginal densities: $f_1(x_1) = \int f_{12}(x_1, u_2)du_2$,
Conditional densities: $f_{1\cdot 2}(x_1 \mid x_2) = \frac{f_{12}(x_1, x_2)}{f_2(x_2)}$
Copula, Sklar's theorem: $f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2); \theta]$
Covariance: $\sigma_{12} = \int_1 \int_2 [x_1 - \mu_1][x_2 - \mu_2] f_{12} dx_1 dx_2$
Correlation: $\rho_{12} = \sigma_{12}/(\sigma_1 \sigma_2)$
Linear transformation of random variables: $E(a + bX) = a + bE(X)$,
$V(a + bX) = b^2 V(X),\ \sigma(a + bX) = b\sigma(X)$
Sum of random variables: $E(X_1 + X_2) = \mu_1 + \mu_2$,
$V(X_1 + X_2) = \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}$
Portfolios of random variables: $Y = w'X,\ E(Y) = \mu_p = w'\mu,\ \sigma_p^2 = w'\Sigma w$
Product of random variables: $E(X_1\ X_2) = \mu_1 \mu_2 + \sigma_{12}$,
$V(X_1\ X_2) = \mu_1^2 \sigma_2^2 + \sigma_1^2 \mu_2^2 + \sigma_1^2 \sigma_2^2$
Uniform distribution: $E(X) = \frac{a+b}{2},\ V(X) = \frac{(b-a)^2}{12}$
Normal distribution: $E(X) = \mu,\ V(X) = \sigma^2,\ \gamma = 0,\ \delta = 3$
Lognormal distribution: for X if $Y = \ln(X)$ is normal, $E[X] = \exp\left[\mu + \frac{1}{2}\sigma^2\right]$,
$V[X] = \exp[2\mu + 2\sigma^2] - \exp[2\mu + \sigma^2]$
Student's t distribution: $V[X] = \frac{k}{k-2},\ \gamma = 0,\ \delta = 3 + \frac{6}{k-4}$
Binomial distribution: $E[X] = pn,\ V[X] = p(1 - p)n$
Poisson distribution: $E[X] = \lambda,\ V[X] = \lambda$
Distribution of averages (CLT): $\bar{X} \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$
Distribution of tails (EVT): $y = (x - u)/\beta \rightarrow$ Generalized Pareto distribution

2.7 ANSWERS TO CHAPTER EXAMPLES

Example 2.1: FRM Exam 2000—Question 81

b. Correlation is a measure of linear association. Independence implies zero correlation, but the reverse is not always true.

Example 2.2: FRM Exam 2007—Question 93

b. The function $x \times y$ is described in the following table. The sum of the entries is 36. The scaling factor k must be such that the total probability is one. Therefore, we have $k = 1/36$. The table shows one instance where $x + y > 5$, which is $x = 3$, $y = 3$. The probability is $p = 9/36 = 1/4$.

$x \times y$	$x = 1$	2	3
$y = 1$	1	2	3
2	2	4	6
3	3	6	9

Example 2.3: FRM Exam 2007—Question 127

b. The variance is $V(3A + 2B) = 3^2V(A) + 2^2V(B) + 2 \times 3\ 2\ \text{Cov}(A, B) = 9\ 1 + 4\ 1 + 12 \times 0.35 = 17.2$.

Example 2.4: FRM Exam 2002—Question 70

b. Statement a. is correct, as it is a linear operation. Statement c. is correct, as in Equation (2.32). Statement d. is correct, as the covariance term is zero if the variables are uncorrelated. Statement b. is false, as adding a constant c to a variable cannot change the variance. The constant drops out because it is also in the expectation.

Example 2.5: FRM Exam 2002—Question 119

a. The mean is the center of the distribution, which is the average of a and b.

Example 2.6: FRM Exam 2005—Question 62

c. The event is the sum of the probabilities $P(-0.5 < Z < +0.5)$ and $P(Z > +1.5)$. Given the symmetry of the normal distribution, or that $N(d) = 1 - N(-d)$, this gives $P(-0.5 < Z < +0.5) = 2P(0 < Z < +0.5) = 2(P(Z < +0.5) - 0.5) = 2(N(0.5) - 0.5) = 2(0.6915 - 0.5) = 0.3830$ and $P(Z > +1.5) = N(-1.5) = 0.0668$. The sum is 0.4498.

Example 2.7: FRM Exam 2003—Question 21

b. Skewness is 0, kurtosis 3, the entire distribution is described by μ and σ, and the p.d.f. is correct.

Example 2.8: FRM Exam 2006—Question 11

d. A platykurtic distribution has kurtosis less than 3, less than the normal p.d.f. because all other answers have higher kurtosis, this produces the lowest extreme values.

Example 2.9: FRM Exam 1999—Question 5

c. X is said to be lognormally distributed if its logarithm $Y = \ln(X)$ is normally distributed.

Example 2.10: FRM Exam 2007—Question 21

a. A lognormal distribution is skewed to the right. Intuitively, if this represents the distribution of prices, prices can fall at most by 100% but can increase by more than that.

Example 2.11: FRM Exam 2002—Question 125

c. Note that this is a two-tailed confidence band, so that $\alpha = 1.96$. We find the extreme values from $\$100\exp(\mu \pm \alpha\sigma)$. The lower limit is then $V_1 = \$100\exp(0.10 - 1.96 \times 0.2) = \$100\exp(-0.292) = \$74.68$. The upper limit is $V_2 = \$100\exp(0.10 + 1.96 \times 0.2) = \$100\exp(0.492) = \$163.56$.

Example 2.12: FRM Exam 2000—Question 128

c. Using Equation (2.48), we have $E[X] = \exp[\mu + 0.5\sigma^2] = \exp[0 + 0.5 * 0.5^2] = 1.1331$. Assuming there is no error in the answers listed for the variance, it is sufficient to find the correct answer for the expected value.

Example 2.13: FRM Exam 2003—Question 18

c. The two distributions have the same skewness of zero but the Student's t has higher kurtosis. As the number of degrees of freedom increases, the Student converges to the normal, so c. is the correct answer.

Example 2.14: FRM Exam 2006—Question 84

d. We use the density given by Equation (2.56). The number of trials is $n = 6$. The probability of guessing correctly just by chance is $p = 1/4 = 0.25$. The probability of zero lucky guesses is $\binom{6}{0}0.25^0 0.75^6 = 0.75^6 = 0.17798$. The probability of one lucky guess is $\binom{6}{1}0.25^1 0.75^5 = 6\ 0.25\ 0.75^5 = 0.35596$. The sum is 0.5339.

Note that the same analysis can be applied to the distribution of scores on the FRM examination with 140 questions. It would be virtually impossible to have a score of zero, assuming random guesses; this probability is $0.75^{140} = 3.2E - 18$ Also, the expected percentage score under random guesses is $p = 25\%$.

Example 2.15: FRM Exam 2004—Question 60

c. The normal approximation to the Poisson improves when the success rate, λ is very high. Because this is also the mean and variance, answers a. and b. are wrong. In turn, the binomial density is well approximated by the Poisson density when $np = \lambda$ is large.

Example 2.16: FRM Exam 2007—Question 110

d. EVT only uses information in the tail, so statement a. is correct. Conventional approaches such as delta-normal VAR assume a fixed p.d.f. for the entire distribution, which may understate the extent of fat tails. So, statement b. is correct. The first step in EVT is to choose a cutoff point for the tail, then to estimate the parameters of the tail distribution, so statement c. is correct. Finally, EVT does not ignore extreme events (as long as they are in the sample).

APPENDIX: REVIEW OF MATRIX MULTIPLICATION

This appendix briefly reviews the mathematics of matrix multiplication. Say that we have two matrices, A and B that we wish to multiply to obtain the new matrix $C = AB$. The respective dimensions are $(n \times m)$ for A, that is, n rows and m columns, and $(m \times p)$ for B. The number of columns for A must exactly match (or conform) to the number of rows for B. If so, this will result in a matrix C of dimensions $(n \times p)$.

We can write the matrix A in terms of its individual components a_{ij}, where i denotes the row and j denotes the column:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

As an illustration, take a simple example where the matrices are of dimension (2×3) and (3×2).

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$C = AB = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

To multiply the matrices, each row of A is multiplied element-by-element by each column of B. For instance, c_{12} is obtained by taking

$$c_{12} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}$$

The matrix C is then:

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

Matrix multiplication can be easily implemented in Excel using the function MMULT. First, we highlight the cells representing the output matrix C, say f1:g2. Then we enter the function, for instance MMULT(a1:c2; d1:e3), where the first range represents the first matrix A, here 2 by 3, and the second range represents the matrix B, here 3 by 2. The final step is to hit the three keys Control-Shift-Return simultaneously.

CHAPTER 3

Fundamentals of Statistics

The preceding chapter was mainly concerned with the theory of probability, including distribution theory. In practice, researchers have to find methods to choose among distributions and to estimate distribution parameters from real data. The subject of sampling brings us now to the theory of statistics. Whereas probability assumes the distributions are known, statistics attempts to make inferences from actual data.

Here, we sample from the distribution of a population, say the change in the exchange rate, to make inferences about the population. The questions are, what is the best distribution for this random variable and what are the best parameters for this distribution? Risk measurement, however, typically deals with large numbers of random variables. So, we also want to characterize the relationships between the risk factors to which the portfolio is exposed. For example, do we observe that movements in the yen/dollar rate are correlated with the dollar/euro rate? Another type of problem is to develop decision rules to test some hypotheses, for instance whether the volatility remains stable over time.

These examples illustrate two important problems in statistical inference, i.e., **estimation** and **tests of hypotheses**. With estimation, we wish to estimate the value of an unknown parameter from sample data. With tests of hypotheses, we wish to verify a conjecture about the data.

This chapter reviews the fundamental tools of statistics theory for risk managers. Section 3.1 discusses the sampling of real data and the construction of returns. The problem of parameter estimation is presented in Section 3.2. Section 3.3 then turns to regression analysis, summarizing important results as well as common pitfalls in their interpretation.

3.1 REAL DATA

To start with an example, let us say that we observe movements in the daily yen/dollar exchange rate and wish to characterize the distribution of tomorrow's exchange rate.

The risk manager's job is to assess the range of potential gains and losses on a trader's position. He or she observes a sequence of past spot prices $S_0, S_1, \ldots, S_t$, from which we have to infer the distribution of tomorrow's price, S_{t+1}.

3.1.1 Measuring Returns

The truly random component in tomorrow's price is not its level, but rather its change relative to today's price. We measure the *relative rate of change* in the spot price:

$$r_t = (S_t - S_{t-1})/S_{t-1} \tag{3.1}$$

Alternatively, we could construct the logarithm of the price ratio:

$$R_t = \ln[S_t/S_{t-1}] \tag{3.2}$$

which is equivalent to using continuous instead of discrete compounding. This is also

$$R_t = \ln[1 + (S_t - S_{t-1})/S_{t-1}] = \ln[1 + r_t]$$

Because $\ln(1 + x)$ is close to x if x is small, R_t should be close to r_t provided the return is small. For daily data, there is typically little difference between R_t and r_t.

The return defined so far is the **capital appreciation return**, which ignores the income payment on the asset. Define the dividend or coupon as D_t. In the case of an exchange rate position, this is the interest payment in the foreign currency over the holding period. The **total return** on the asset is

$$r_t^{\text{TOT}} = (S_t + D_t - S_{t-1})/S_{t-1} \tag{3.3}$$

When the horizon is very short, the income return is typically very small compared to the capital appreciation return.

The next question is whether the sequence of variables r_t can be viewed as independent observations. If so, one could hypothesize, for instance, that the random variables are drawn from a normal distribution $N(\mu, \sigma^2)$. We could then proceed to estimate μ and σ^2 from the data and use this information to create a distribution for tomorrow's spot price change.

Independent observations have the very nice property that their joint distribution is the product of their marginal distribution, which considerably simplifies the analysis. The obvious question is whether this assumption is a workable approximation. In fact, there are good economic reasons to believe that rates of change on financial prices are close to independent.

The hypothesis of **efficient markets** postulates that current prices convey all relevant information about the asset. If so, any change in the asset price must be due to news, or events which are by definition impossible to forecast (otherwise, it would not be news). This implies that changes in prices are unpredictable and, hence, satisfy our definition of independent random variables.

This hypothesis, also known as the **random walk** theory, implies that the conditional distribution of returns depends only on current prices, and not on the previous history of prices. If so, technical analysis must be a fruitless exercise. Technical analysts try to forecast price movements from past price patterns.

If in addition the distribution of returns is constant over time, the variables are said to be **independently and identically distributed** (i.i.d.). So, we could consider that the observations r_t are independent draws from the same distribution $N(\mu, \sigma^2)$.

Later, we will consider deviations from this basic model. Distributions of financial returns typically display fat tails. Also, variances are not constant and display some persistence; expected returns can also slightly vary over time.

3.1.2 Time Aggregation

It is often necessary to translate parameters over a given horizon to another horizon. For example, we may have raw data for daily returns, from which we compute a daily volatility that we want to extend to a monthly volatility.

Returns can be easily related across time when we use the log of the price ratio, because the log of a product is the sum of the logs of the individual terms. The two-day return, for example, can be decomposed as

$$R_{02} = \ln[S_2/S_0] = \ln[(S_2/S_1) \times (S_1/S_0)] = \ln[S_1/S_0] + \ln[S_2/S_1] = R_{01} + R_{12} \tag{3.4}$$

This decomposition is only approximate if we use discrete returns, however.

The expected return and variance are then $E(R_{02}) = E(R_{01}) + E(R_{12})$ and $V(R_{02}) = V(R_{01}) + V(R_{12}) + 2\text{Cov}(R_{01}, R_{12})$. Assuming returns are uncorrelated and have identical distributions across days, we have $E(R_{02}) = 2E(R_{01})$ and $V(R_{02}) = 2V(R_{01})$.

Generalizing over T days, we can relate the moments of the T-day returns R_T to those of the 1-day returns R_1:

$$E(R_T) = E(R_1)T \tag{3.5}$$

$$V(R_T) = V(R_1)T \tag{3.6}$$

Expressed in terms of volatility, this yields the **square root of time rule**:

$$\text{SD}(R_T) = \text{SD}(R_1)\sqrt{T} \tag{3.7}$$

KEY CONCEPT

When successive returns are uncorrelated, the volatility increases as the horizon extends following the square root of time.

More generally, the variance can be added up from different values across different periods. For instance, the variance over the next year can be computed as the average monthly variance over the first three months, multiplied by 3, plus the average variance over the last nine months, multiplied by 9. This type of analysis is routinely used to construct a term structure of implied volatilities, which are derived from option data for different maturities.

It should be emphasized that this holds only if returns have constant parameters across time and are uncorrelated. When there is non-zero correlation across days, the two-day variance is

$$V(R_2) = V(R_1) + V(R_1) + 2\rho V(R_1) = 2V(R_1)(1+\rho) \quad (3.8)$$

Because we are considering correlations in the time series of the same variable, ρ is called the **autocorrelation coefficient**, or the **serial autocorrelation coefficient.** A positive value for ρ implies that a movement in one direction in one day is likely to be followed by another movement in the same direction the next day. A positive autocorrelation signals the existence of a **trend.** In this case, Equation (3.8) shows that the two-day variance is greater than the one obtained by the square root of time rule.

A negative value for ρ implies that a movement in one direction in one day is likely to be followed by a movement in the other direction the next day. So, prices tend to revert back to a mean value. A negative autocorrelation signals

EXAMPLE 3.1: FRM EXAM 1999—QUESTION 4

A fundamental assumption of the random walk hypothesis of market returns is that returns from one time period to the next are statistically independent. This assumption implies

a. Returns from one time period to the next can never be equal.
b. Returns from one time period to the next are uncorrelated.
c. Knowledge of the returns from one time period does not help in predicting returns from the next time period.
d. Both b) and c) are true.

EXAMPLE 3.2: FRM EXAM 2002—QUESTION 3

Consider a stock with daily returns that follow a random walk. The annualized volatility is 34%. Estimate the weekly volatility of this stock assuming that the year has 52 weeks.

a. 6.80%
b. 5.83%
c. 4.85%
d. 4.71%

EXAMPLE 3.3: FRM EXAM 2002—QUESTION 2

Assume we calculate a one-week VAR for a natural gas position by rescaling the daily VAR using the square-root rule. Let us now assume that we determine the *true* gas price process to be mean-reverting and recalculate the VAR.

Which of the following statements is true?

a. The recalculated VAR will be less than the original VAR.
b. The recalculated VAR will be equal to the original VAR.
c. The recalculated VAR will be greater than the original VAR.
d. There is no necessary relation between the recalculated VAR and the original VAR.

mean reversion. In this case, the two-day variance is less than the one obtained by the square root of time rule.

3.1.3 Portfolio Aggregation

Let us now turn to aggregation of returns across assets. Consider, for example, an equity portfolio consisting of investments in N shares. Define the number of each share held as q_i with unit price S_i. The portfolio value at time t is then

$$W_t = \sum_{i=1}^{N} q_i S_{i,t} \tag{3.9}$$

We can write the weight assigned to asset i as

$$w_{i,t} = \frac{q_i S_{i,t}}{W_t} \tag{3.10}$$

which by construction sum to unity. Using weights, however, rules out situations with zero net investment, $W_t = 0$, such as some derivatives positions. But we could have positive and negative weights if short selling is allowed, or weights greater than one if the portfolio can be leveraged.

The next period, the portfolio value is

$$W_{t+1} = \sum_{i=1}^{N} q_i S_{i,t+1} \tag{3.11}$$

assuming that the unit price incorporates any income payment. The gross, or dollar, return is then

$$W_{t+1} - W_t = \sum_{i=1}^{N} q_i (S_{i,t+1} - S_{i,t}) \tag{3.12}$$

and the *rate* of return is

$$\frac{W_{t+1} - W_t}{W_t} = \sum_{i=1}^{N} \frac{q_i S_{i,t}}{W_t} \frac{(S_{i,t+1} - S_{i,t})}{S_{i,t}} = \sum_{i=1}^{N} w_{i,t} \frac{(S_{i,t+1} - S_{i,t})}{S_{i,t}} \tag{3.13}$$

So, the portfolio rate of return is a linear combination of the asset returns

$$r_{p,t+1} = \sum_{i=1}^{N} w_{i,t}\, r_{i,t+1} \tag{3.14}$$

The dollar return is then

$$W_{t+1} - W_t = \left[\sum_{i=1}^{N} w_{i,t} r_{i,t+1} \right] W_t \tag{3.15}$$

and has a normal distribution if the individual returns are also normally distributed.

Alternatively, we could express the individual positions in dollar terms,

$$x_{i,t} = w_{i,t} W_t = q_i S_{i,t} \tag{3.16}$$

The dollar return is also, using dollar amounts,

$$W_{t+1} - W_t = \left[\sum_{i=1}^{N} x_{i,t} r_{i,t+1} \right] \tag{3.17}$$

As we have seen in the previous chapter, the variance of the portfolio dollar return is

$$V[W_{t+1} - W_t] = x' \Sigma x \tag{3.18}$$

Because the portfolio follows a normal distribution, it is fully characterized by it expected return and variance. The portfolio VAR is then

$$\text{VAR} = \alpha \sqrt{x' \Sigma x} \tag{3.19}$$

where α depends on the selected density function.

EXAMPLE 3.4: FRM EXAM 2004—QUESTION 39

Consider a portfolio with 40% invested in asset X and 60% invested in asset Y. The mean and variance of return on X are 0 and 25, respectively. The mean and variance of return on Y are 1 and 121, respectively. The correlation coefficient between X and Y is 0.3. What is the nearest value for portfolio volatility?

a. 9.51
b. 8.60
c. 13.38
d. 7.45

3.2 PARAMETER ESTIMATION

Armed with our i.i.d. sample of T observations, we can start estimating the parameters of interest, such as the sample mean, the variance, and other moments.

3.2.1 Distribution of Estimates

As in the previous chapter, define x_i as the realization of a random sample. The expected return, or mean, $\mu = E(X)$ can be estimated by the sample mean,

$$m = \widehat{\mu} = \frac{1}{T}\sum_{i=1}^{T} x_i \tag{3.20}$$

Intuitively, we assign the same weight of $1/T$ to all observations because they all have the same probability. The variance, $\sigma^2 = E[(X-\mu)^2]$, can be estimated by the sample variance,

$$s^2 = \widehat{\sigma}^2 = \frac{1}{(T-1)}\sum_{i=1}^{T}(x_i - \widehat{\mu})^2 \tag{3.21}$$

Note that we divide by $T-1$ instead of T. This is because we estimate the variance around an unknown parameter, the mean. So, we have fewer degrees of freedom than otherwise. As a result, we need to adjust s^2 to ensure that its expectation equals the true value. In most situations, however, T is large so that this adjustment is minor.

It is essential to note that these estimated values depend on the particular sample and, hence, have some inherent variability. The sample mean itself is distributed as

$$m = \widehat{\mu} \sim N(\mu, \sigma^2/T) \tag{3.22}$$

If the population distribution is normal, this exactly describes the distribution of the sample mean. Otherwise, the central limit theorem states that this distribution is only valid asymptotically, i.e., for large samples.

For the distribution of the sample variance $\widehat{\sigma}^2$, one can show that, when X is normal, the following ratio is distributed as a chi-square with $(T-1)$ degrees of freedom

$$\frac{(T-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(T-1) \tag{3.23}$$

If the sample size T is large enough, the chi-square distribution converges to a normal distribution:

$$\widehat{\sigma}^2 \sim N\left(\sigma^2, \sigma^4 \frac{2}{(T-1)}\right) \tag{3.24}$$

Using the same approximation, the sample standard deviation has a normal distribution with a standard error of

$$\mathrm{se}(\widehat{\sigma}) = \sigma\sqrt{\frac{1}{2T}} \tag{3.25}$$

We can use this information for **hypothesis testing**. For instance, we would like to detect a constant trend in X. Here, the **null hypothesis** is that $\mu = 0$. To answer the question, we use the distributional assumption in Equation (3.22) and compute a standard normal variable as the ratio of the estimated mean to its standard error, or

$$z = \frac{(m-0)}{\sigma/\sqrt{T}} \tag{3.26}$$

Because this is now a standard normal variable, we would not expect to observe values far away from zero. Typically, we would set the significance level at 95 percent, which translates into a two-tailed interval for z of $[-1.96, +1.96]$. Roughly, this means that, if the absolute value of z is greater than two, we would reject the hypothesis that m came from a distribution with a mean of zero. We can have some confidence that the true μ is indeed different from zero.

In fact, we do not know the true σ and use the estimated s instead. The distribution is a Student's t with T degrees of freedom:

$$t = \frac{(m-0)}{s/\sqrt{T}} \tag{3.27}$$

for which the cutoff values can be found from tables. For large values of T, however, this distribution is close to the normal.

Example

We want to characterize movements in the monthly yen/dollar exchange rate from historical data, taken over 1990 to 1999. Returns are defined in terms of continuously compounded changes, as in Equation (3.2). The sample size is $T = 120$, and estimated parameters are $m = -0.28\%$ and $s = 3.55\%$ (per month).

Using Equation (3.22), the standard error of the mean is approximately $\text{se}(m) = s/\sqrt{T} = 0.32\%$. For the null of $\mu = 0$, this gives a t-ratio of $t = m/\text{se}(m) = -0.28\%/0.32\% = -0.87$. Because this number is less than 2 in absolute value, we cannot reject the hypothesis that the mean is zero at the 95% confidence level. This is a typical result for financial series. The mean is not precisely estimated.

Next, we turn to the precision in the sample standard deviation. By Equation (3.25), its standard error is $\text{se}(s) = \sigma\sqrt{\frac{1}{(2T)}} = 0.229\%$. For the null of $\sigma = 0$, this gives a ratio of $z = s/\text{se}(s) = 3.55\%/0.229\% = 15.5$, which is very high. So, the volatility is not zero. Therefore, there is much more precision in the measurement of s than in that of m.

Furthermore, we can construct 95% confidence intervals around the estimated values. These are:

$$[m - 1.96 \times \text{se}(m), m + 1.96 \times \text{se}(m)] = [-0.92\%, +0.35\%]$$
$$[s - 1.96 \times \text{se}(s), s + 1.96 \times \text{se}(s)] = [3.10\%, 4.00\%]$$

So, we could be reasonably confident that the volatility is between 3% and 4%, but we cannot even be sure that the mean is different from zero.

3.2.2 Choosing Significance Levels for Tests

Hypothesis testing requires the choice of a significance level, which needs careful consideration. Two types of errors can arise, as described in Table 3.1. A *type 1* error involves rejecting a correct model. A *type 2* error involves accepting an incorrect model. For a given test, increasing the significance level will decrease the probability of a type 1 error but increase the probability of a type 2 error. Thus, the choice of the significance level should reflect the cost of each of these errors.

This type of situation arises, for example, when a risk manager or regulator must decide whether to accept a VAR model. The first step is to record the number of exceptions, or losses worse than VAR forecasts constructed at the 99% level

TABLE 3.1 Decision Errors

	Model	
Decision:	Correct	Incorrect
Accept	OK	Type 2 error
Reject	Type 1 error	OK

of confidence, over the last 250 days, for example. Under the null hypothesis that the VAR model is correctly calibrated, the number of exceptions should follow a binomial distribution with expected value of $E[X] = np = 250(1 - 0.99) = 2.5$. The risk manager then has to pick a cutoff number of exceptions above which the model would be rejected. The type 1 error rate is the probability of observing higher numbers than the cutoff point. Say the risk manager chooses $n = 4$, which corresponds to a type 1 error rate or significance level of 10.8%. Once this cutoff point is selected, however, it could lead the risk manager to incorrectly accept a VAR model that has a lower confidence level, say 97%.

Suppose for instance that the risk manager observes six exceptions. The **p-value** of observing six or more exceptions is 4.1%. Because this is below the selected significance level, the risk manager would have to conclude that the VAR model is incorrect.

3.2.3 Precision of Estimates

Equation (3.22) shows that, when the sample size increases, the standard error of $\widehat{\mu}$ shrinks at a rate proportional to $1/\sqrt{T}$. The precision of the estimate increases as the number of observations increases.

This result will prove useful to assess the precision of estimates generated from **numerical simulations**, which are widely used in risk management. Numerical simulations create independent random variables over a fixed number of replications T. If T is too small, the final estimates will be imprecisely measured. If T is very large, the estimates will be accurate. The precision of the estimates increases at a rate proportional to $1/\sqrt{T}$.

KEY CONCEPT

With independent draws, the standard deviation of most statistics is inversely related to the square root of number of observations T. Thus, more observations make for more precise estimates.

EXAMPLE 3.5: FRM EXAM 2007—QUESTION 137

What does a hypothesis test at the 5% significance level mean?

a. P(not reject H0 | H0 is true) = 0.05
b. P(not reject H0 | H0 is false) = 0.05
c. P(reject H0 | H0 is true) = 0.05
d. P(reject H0 | H0 is false) = 0.05

EXAMPLE 3.6: FRM EXAM 2007—QUESTION 2

Which of the following statements regarding hypothesis testing is *incorrect*?

a. Type II error refers to the failure to reject the null hypothesis when it is actually false.
b. Hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from that population.
c. All else being equal, the decrease in the chance of making a type I error comes at the cost of increasing the probability of making a type II error.
d. The p-value decision rule is to reject the null hypothesis if the p-value is greater than the significance level.

3.3 REGRESSION ANALYSIS

Regression analysis has particular importance for risk management, because it can be used to explain and forecast financial variables.

3.3.1 Bivariate Regression

In a **linear regression**, the **dependent variable** y is projected on a set of N predetermined **independent variables**, x. In the simplest bivariate case we write

$$y_t = \alpha + \beta x_t + \epsilon_t, \qquad t = 1, \ldots, T \tag{3.28}$$

where α is called the **intercept**, or constant, β is called the **slope**, and ϵ is called the **residual**, or **error term**. This could represent a time-series or a cross-section.

The **ordinary least squares** (OLS) assumptions are

- *The errors are independent of x.*
- *The errors have a normal distribution with zero mean and constant variance, conditional on x.*
- *The errors are independent across observations.*

Based on these assumptions, the usual methodology is to estimate the coefficients by minimizing the sum of squared errors. Beta is estimated by

$$\hat{\beta} = \frac{[1/(T-1)]\sum_t (x_t - \bar{x})(y_t - \bar{y})}{[1/(T-1)]\sum_t (x_t - \bar{x})^2} \tag{3.29}$$

where $\bar{x}$ and $\bar{y}$ correspond to the means of x_t and y_t. Alpha is estimated by

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x} \tag{3.30}$$

Note that the numerator in Equation (3.29) is also the sample covariance between two series x_i and x_j, which can be written as

$$\widehat{\sigma_{ij}} = \frac{1}{(T-1)}\sum_{t=1}^{T}(x_{t,i} - \widehat{\mu_i})(x_{t,j} - \widehat{\mu_j}) \tag{3.31}$$

To interpret β, we can take the covariance between y and x, which is

$$\text{Cov}(y, x) = \text{Cov}(\alpha + \beta x + \epsilon, x) = \beta\text{Cov}(x, x) = \beta V(x)$$

because ϵ is conditionally independent of x. This shows that the population β is also

$$\beta(y, x) = \frac{\text{Cov}(y, x)}{V(x)} = \frac{\rho(y, x)\sigma(y)\sigma(x)}{\sigma^2(x)} = \rho(y, x)\frac{\sigma(y)}{\sigma(x)} \tag{3.32}$$

The **regression fit** can be assessed by examining the size of the residuals, obtained by subtracting the fitted values $\widehat{y_t}$ from y_t,

$$\widehat{\epsilon_t} = y_t - \widehat{y_t} = y_t - \widehat{\alpha} - \widehat{\beta}x_t \tag{3.33}$$

and taking the estimated variance as

$$V(\widehat{\epsilon}) = \frac{1}{(T-2)}\sum_{t=1}^{T}\widehat{\epsilon_t}^2 \tag{3.34}$$

We divide by $T - 2$ because the estimator uses two unknown quantities, $\widehat{\alpha}$ and $\widehat{\beta}$. Also note that, because the regression includes an intercept, the average value of $\widehat{\epsilon}$ has to be exactly zero.

The quality of the fit can be assessed using a unitless measure called the **regression R-square,** also called **coefficient of determination.** This is defined as

$$R^2 = 1 - \frac{\text{SSE}}{\text{SSY}} = 1 - \frac{\sum_t \widehat{\epsilon_t}^2}{\sum_t (y_t - \bar{y})^2} \tag{3.35}$$

where SSE is the sum of squared errors, and SSY is the sum of squared deviations of y around its mean. If the regression includes a constant, we always have $0 \leq R^2 \leq 1$. In this case, R-square is also the square of the usual correlation coefficient,

$$R^2 = \rho(y, x)^2 \tag{3.36}$$

The R^2 measures the degree to which the size of the errors is smaller than that of the original dependent variables y. To interpret R^2, consider two extreme cases.

On one hand, if the fit is excellent, all the errors will be zero, and the numerator in Equation (3.35) will be zero, which gives $R^2 = 1$. On the other hand, if the fit is poor, SSE will be as large as SSY and the ratio will be one, giving $R^2 = 0$.

Alternatively, we can interpret the R-square by decomposing the variance of $y_t = \alpha + \beta x_t + \epsilon_t$. Because ϵ and x are uncorrelated, this yields

$$V(y) = \beta^2 V(x) + V(\epsilon) \tag{3.37}$$

Dividing by $V(y)$,

$$1 = \frac{\beta^2 V(x)}{V(y)} + \frac{V(\epsilon)}{V(y)} \tag{3.38}$$

Because the R-square is also $R^2 = 1 - V(\epsilon)/V(y)$, it is equal to $= \beta^2 V(x)/V(y)$, which is the contribution in the variation of y due to β and x.

Finally, we can derive the distribution of the estimated coefficients, which is normal and centered around the true values. For the slope coefficient, $\widehat{\beta} \sim N(\beta, V(\widehat{\beta}))$, with variance given by

$$V(\widehat{\beta}) = V(\widehat{\epsilon}) \frac{1}{\sum_t (x_t - \bar{x})^2} \tag{3.39}$$

This can be used to test whether the slope coefficient is significantly different from zero. The associated test statistic

$$t = \widehat{\beta}/\sigma(\widehat{\beta}) \tag{3.40}$$

has a Student's t distribution. Typically, if the absolute value of the statistic is above 2, we would reject the hypothesis that there is no relationship between y and x. This corresponds to a two-tailed significance level of 5%.

3.3.2 Autoregression

A particularly useful application is a regression of a variable on a lagged value of itself, called **autoregression**

$$y_t = \alpha + \beta_k y_{t-k} + \epsilon_t, \qquad t = 1, \ldots, T \tag{3.41}$$

If the β coefficient is significant, previous movements in the variable can be used to predict future movements. Here, the coefficient β_k is known as the kth-order **autocorrelation coefficient.**

Consider for instance a first-order autoregression, where the daily change in the yen/dollar rate is regressed on the previous day's value. A positive coefficient $\widehat{\beta_1}$

indicates a trend. A negative coefficient indicates mean reversion. As an example, assume that we find that $\widehat{\beta_1} = 0.10$, with zero intercept. One day, the yen goes up by 2%. Our best forecast for the next day is another upmove of

$$E[y_{t+1}] = \beta_1 y_t = 0.1 \times 2\% = 0.2\%$$

Autocorrelation changes normal patterns in risk across horizons. When there is no autocorrelation, risk increases with the square root of time. With positive autocorrelation, shocks have a longer-lasting effect and risk increases faster than the square root of time.

3.3.3 Multivariate Regression

More generally, the regression in Equation (3.28) can be written, with N independent variables:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1N} \\ \vdots & & & & \\ x_{T1} & x_{T2} & x_{T3} & \dots & x_{TN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix} \tag{3.42}$$

This can include the case of a constant when the first column of X is a vector of ones, in which case β_1 is the usual α. In matrix notation,

$$y = X\beta + \epsilon \tag{3.43}$$

The estimated coefficients can be written in matrix notation as

$$\widehat{\beta} = (X'X)^{-1}X'y \tag{3.44}$$

and their covariance matrix as

$$V(\widehat{\beta}) = \sigma^2(\epsilon)(X'X)^{-1} \tag{3.45}$$

We can extend the t-statistic to a multivariate environment. Say we want to test whether the last m coefficients are jointly zero. Define $\widehat{\beta_m}$ as these grouped coefficients and $V_m(\widehat{\beta})$ as their covariance matrix. We set up a statistic

$$F = \frac{\widehat{\beta}'_m V_m(\widehat{\beta})^{-1}\widehat{\beta}_m/m}{\text{SSE}/(T-N)} \tag{3.46}$$

which has an F-distribution with m and $T-N$ degrees of freedom. As before, we would reject the hypothesis if the value of F is too large compared to critical values from tables.

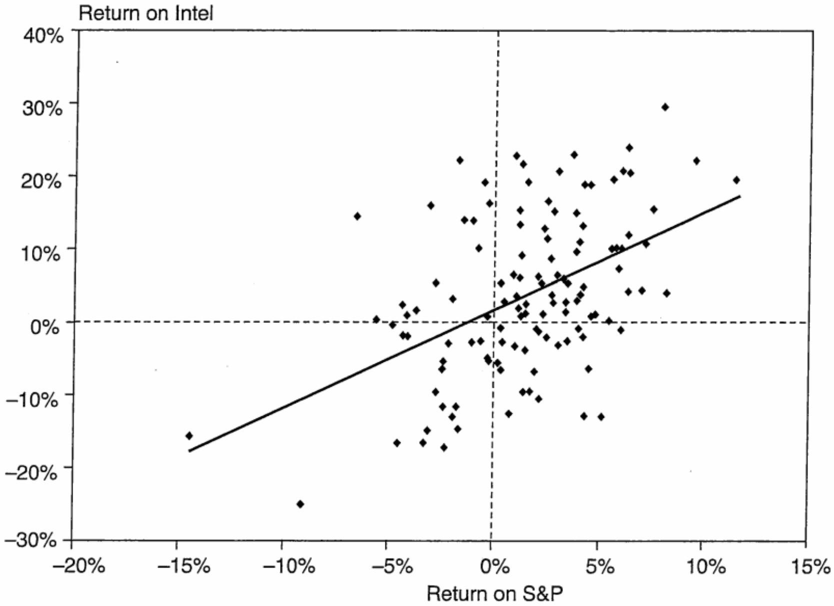

FIGURE 3.1 Intel Return vs. S&P Return

3.3.4 Example

This section gives the example of a regression of a stock return on the market. Such analysis is commonly used to assess whether movements in the stock can be hedged using stock-market index futures.

We consider 10 years of data for Intel and the S&P 500, using total rates of return over a month. Figure 3.1 plots the 120 combination of returns, or (y_t, x_t). Apparently, there is a positive relationship between the two variables, as shown by the straight line that represents the regression fit $(\widehat{y}_t, x_t)$.

Table 3.2 displays the regression results. The regression shows a positive relationship between the two variables, with $\widehat{\beta} = 1.349$. This is significantly positive, with a standard error of 0.229 and t-statistic of 5.90. The t-statistic is very high, with an associated probability value (p-value) close to zero. Thus, we can be fairly confident of a positive association between the two variables.

This beta coefficient is also called **systematic risk**, or exposure to general market movements. Typically, technology stocks have greater systematic risk than

TABLE 3.2 Regression Results
$y = \alpha + \beta x$, $y =$ Intel return, $x =$ S&P return

R-square	0.228			
Standard error of y	10.94%			
Standard error of $\widehat{\epsilon}$	9.62%			
Coefficient	**Estimate**	**Standard Error**	**T-statistic**	**P-value**
Intercept $\widehat{\alpha}$	0.0168	0.0094	1.78	0.77
Intercept $\widehat{\beta}$	1.349	0.229	5.90	0.00

the average. Indeed, the slope in Intel's regression is greater than unity. To test whether β is significantly different from 1, we can compute a z-score as

$$z = \frac{(\widehat{\beta} - 1)}{s(\widehat{\beta})} = \frac{(1.349 - 1)}{0.229} = 1.53$$

This is less than the usual cutoff value of 2, so we cannot say for certain that Intel's systematic risk is greater than one.

The R-square of 22.8% can be also interpreted by examining the reduction in dispersion from y to $\widehat{\epsilon}$, which is from 10.94% to 9.62%. The R-square can be written as

$$R^2 = 1 - \frac{9.62\%^2}{10.94\%^2} = 22.8\%$$

Thus, about 23% of the variance of Intel's returns can be attributed to the market.

EXAMPLE 3.7: FRM EXAM 2004—QUESTION 4

Consider the following linear regression model: $Y = a + b\ X + e$. Suppose $a = 0.05$, $b = 1.2$, $\mathrm{SD}(Y) = 0.26$, $\mathrm{SD}(e) = 0.1$, what is the correlation between X and Y?

a. 0.923
b. 0.852
c. 0.701
d. 0.462

EXAMPLE 3.8: FRM EXAM 2007—QUESTION 22

Consider two stocks, A and B. Assume their annual returns are jointly normally distributed, the marginal distribution of each stock has mean 2% and standard deviation 10%, and the correlation is 0.9. What is the expected annual return of stock A if the annual return of stock B is 3%?

a. 2%
b. 2.9%
c. 4.7%
d. 1.1%

EXAMPLE 3.9: FRM EXAM 2004—QUESTION 23

Which of the following statements about the linear regression of the return of a portfolio over the return of its benchmark presented below are correct?

Portfolio parameter	Value
Beta	1.25
Alpha	0.26
Coefficient of determination	0.66
Standard deviation of error	2.42

I. The correlation is 0.71.
II. 34% of the variation in the portfolio return is explained by variation in the benchmark return.
III. The portfolio is the dependent variable.
IV. For an estimated portfolio return of 12%, the confidence interval at 95% is (7.16% to −16.84%).

a. II and IV
b. III and IV
c. I, II, and III
d. II, III, and IV

3.3.5 Pitfalls with Regressions

As with any quantitative method, the usefulness of regression analysis depends on the underlying assumptions being fulfilled for the problem at hand. Potential problems of interpretation are now briefly mentioned.

The original OLS setup assumes that the X variables are predetermined (i.e., exogenous or fixed), as in a controlled experiment. In practice, regressions are performed on actual, existing data that do not satisfy these strict conditions. In the previous regression, returns on the S&P are certainly not predetermined.

If the X variables are stochastic, however, most of the OLS results are still valid as long as the X variables are distributed independently of the errors and their distribution does not involve β and σ^2.

Violations of this assumption are serious because they create biases in the slope coefficients. Biases could lead the researcher to come to the wrong conclusion. For instance, we could have measurement errors in the X variables, which causes the measured X to be correlated with ϵ. This so-called **errors in the variables** problem causes a downward bias, or reduces the estimated slope coefficients from their true values. Note that errors in the y variables are not an issue, because they are captured by the error component ϵ.

A related problem is that of **specification error**. Suppose the true model has N variables but we only use a subset N_1. If the omitted variables are correlated with the included variables, the estimated coefficients will be biased. This is a very serious problem because it is difficult to identify. Biases in the coefficients cause problems with **estimation.**

Another class of problems has to do with potential biases in the standard errors of the coefficients. These errors are especially serious if standard errors are underestimated, creating a sense of false precision in the regression results and perhaps leading to the wrong conclusions. The OLS approach assumes that the errors are independent across observations. This is generally the case for financial time series, but often not in cross-sectional setups. For instance, consider a cross-section of mutual fund returns on some attribute. Mutual fund families often have identical funds, except for the fee structure (e.g., called A for a front load, B for a deferred load). These funds, however, are invested in the same securities and have the same manager. Thus, their returns are certainly not independent. If we run a standard OLS regression with all funds, the standard errors will be too small because we overestimate the number of independent observations. More generally, one has to check that there is no systematic correlation pattern in the residuals. Even with time series, problems can arise with **autocorrelation** in the errors. Biases in the standard errors cause problems with **inference**, as one could conclude erroneously that a coefficient is statistically significant.

Problems with **efficiency** arise when the estimation does not use all available information. For instance, the residuals can have different variances across observations, in which case we have **heteroskedasticity**. This is the opposite of the constant variance case, or **homoskedasticity**. Conditional heteroskedasticity occurs when the variance is systematically related to the independent variables. For instance, large values of X could be associated with high error variances. These problems can be identified by diagnostic checks on the residuals. If heteroskedasticity is present, one could construct better standard errors, or try an alternative specification. This is much less of a problem than problems with estimation or inference, however. Inefficient estimates do not necessarily create biases.

Also, regressions may be subject to **multicollinearity**. This arises when the X variables are highly correlated. Some of the variables may be superfluous, for example using two currencies that are fixed to each other. As a result, the matrix $(X'X)$ in Equation (3.44) will be unstable, and the estimated β unreliable. This problem will show up in large standard errors, however. It can be fixed by discarding some of the variables that are highly correlated with others.

Last, even if all the OLS conditions are satisfied, one has to be extremely careful about using a regression for forecasting. Unlike physical systems, which are inherently stable, financial markets are dynamic and relationships can change quickly. Indeed, financial anomalies, which show up as strongly significant coefficients in historical regressions, have an uncanny ability to disappear as soon as one tries to exploit them.

EXAMPLE 3.10: FRM EXAM 2004—QUESTION 59

Which of the following statements regarding linear regression is *false*?

a. Heteroskedasticity occurs when the variance of residuals is not the same across all observations in the sample.
b. Unconditional heteroskedasticity leads to inefficient estimates, whereas conditional heteroskedasticity can lead to problems with both inference and estimation.
c. Serial correlation occurs when the residual terms are correlated with each other.
d. Multicollinearity occurs when a high correlation exists between or among two or more of the independent variables in a multiple regression.

EXAMPLE 3.11: FRM EXAM 1999—QUESTION 2

Under what circumstances could the explanatory power of regression analysis be overstated?

a. The explanatory variables are not correlated with one another.
b. The variance of the error term decreases as the value of the dependent variable increases.
c. The error term is normally distributed.
d. An important explanatory variable is omitted that influences the explanatory variables included, and the dependent variable.

3.4 IMPORTANT FORMULAS

Discrete returns, log returns: $r_t = (S_t - S_{t-1})/S_{t-1}$, $R_t = \ln[S_t/S_{t-1}]$

Time aggregation: $E(R_T) = E(R_1)T$, $V(R_T) = V(R_1)T$, $\mathrm{SD}(R_T) = \mathrm{SD}(R_1)\sqrt{T}$

Portfolio rate of return, variance: $r_{p,t+1} = \sum_{i=1}^{N} w_{i,t}\, r_{i,t+1} = w'R$,
$V[r_{p,t+1}] = w'\Sigma w$

Estimated mean, variance: $m = \widehat{\mu} = \frac{1}{T}\sum_{i=1}^{T} x_i$, $s^2 = \widehat{\sigma}^2 = \frac{1}{(T-1)}\sum_{i=1}^{T}(x_i - \widehat{\mu})^2$

Distribution of estimated mean, variance, standard deviation: $m = \widehat{\mu} \sim N(\mu, \sigma^2/T)$, $\frac{(T-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(T-1)$, $\widehat{\sigma}^2 \rightarrow N\left(\sigma^2, \sigma^4\frac{2}{(T-1)}\right)$, $\mathrm{se}(\widehat{\sigma}) = \sigma\sqrt{\frac{1}{2T}}$

Bivariate, multivariate regression: $y_t = \alpha + \beta x_t + \epsilon_t$, $y = X\beta + \epsilon$

Estimated beta: $\widehat{\beta} = (X'X)^{-1}X'y$

Population beta: $\beta(y, x) = \frac{\text{Cov}(y,x)}{V(x)} = \frac{\rho(y,x)\sigma(y)\sigma(x)}{\sigma^2(x)} = \rho(y, x)\frac{\sigma(y)}{\sigma(x)}$

Regression R-square: $R^2 = 1 - \frac{\text{SSE}}{\text{SSY}} = 1 - \frac{\sum_t \hat{\epsilon}_t^2}{\sum_t (y_t - \bar{y})^2}$

Variance decomposition: $V(y) = \beta^2 V(x) + V(\epsilon)$

T-statistic for hypothesis of zero coefficient: $t = \hat{\beta}/\sigma(\hat{\beta})$

3.5 ANSWERS TO CHAPTER EXAMPLES

Example 3.1: FRM Exam 1999—Question 4

d. Efficient markets implies that the distribution of future returns does not depend on past returns. Hence, returns cannot be correlated. It could happen, however, that return distributions are independent, but that, just by chance, two successive returns are equal.

Example 3.2: FRM Exam 2002—Question 3

d. Assuming a random walk, we can use the square root of time rule. The weekly volatility is then $34\% \times 1/\sqrt{52} = 4.71\%$.

Example 3.3: FRM Exam 2002—Question 2

a. With mean reversion, the volatility grows more slowly than the square root of time

Example 3.4: FRM Exam 2004—Question 39

d. The variance of the portfolio is given by $\sigma_p^2 = (0.4)^2 25 + (0.6)^2 121 + 2(0.4)(0.6)0.3\sqrt{25 \times 121} = 55.48$. Hence, the volatility is 7.45.

Example 3.5: FRM Exam 2007—Question 137

c. The significance level is the probability of committing a type 1 error, or rejecting a correct model. This is also P(reject H0 | H0 is true). On the other hand, the type 2 error rate is P(not reject H0 | H0 is false).

Example 3.6: FRM Exam 2007—Question 2

d. We would reject the null if the observed p-value is *lower* (not greater) than the significance level.

Example 3.7: FRM Exam 2004—Question 4

a. We can find the volatility of X from the variance decomposition, Equation (3.37). This gives $V(x) = [V(y) - V(e)]/\beta^2 = [0.26^2 - 0.10^2]/1.2^2 = 0.04$. Then $\text{SD}(X) = 0.2$, and $\rho = \beta\text{SD}(X)/\text{SD}(Y) = 1.20.2/0.26 = 0.923$.

Example 3.8: FRM Exam 2007—Question 22

b. The information in this question can be used to construct a regression model of A on B. We have $R_A = 2\% + 0.9(10\%/10\%)(R_B - 2\%) + \epsilon$. Next, replacing R_B by 3% gives $\hat{R}_A = 2\% + 0.9(3\% - 2\%) = 2.9\%$.

Example 3.9: FRM Exam 2004—Question 23

b. The correlation is given by $\sqrt{0.66} = 0.81$, so I. is incorrect. Next, 66% of the variation in Y is explained by the benchmark, so answer II. is incorrect. The portfolio return is indeed the dependent variable Y, so answer III. is correct. Finally, to find the 95% two-tailed confidence interval, we use α from a normal distribution, which covers 95% within plus or minus 1.96, close to 2.00. The interval is then $y - 2\text{SD}(e)$, $y + 2\text{SD}(e)$, or $(7.16 - 16.84)$. So answers III. and IV. are correct.

Example 3.10: FRM Exam 2004—Question 59

b. Heteroskedasticity indeed occurs when the variance of the residuals is not constant, so a. is correct. This leads to inefficient estimates but otherwise does not cause problems with inference and estimation. Statements c. and d. are correct.

Example 3.11: FRM Exam 1999—Question 2

d. If the true regression includes a third variable z that influences both y and x, the error term will not be conditionally independent of x, which violates one of the assumptions of the OLS model. This will artificially increase the explanatory power of the regression. Intuitively, the variable x will appear to explain more of the variation in y simply because it is correlated with z.

CHAPTER 4

Monte Carlo Methods

The two preceding chapters dealt with probability and statistics. The former involves the generation of random variables from known distributions. The second deals with estimation of distribution parameters from actual data. With estimated distributions in hand, we can proceed to the next step, which is the simulation of random variables for the purpose of risk management. Such simulations, called **Monte Carlo** simulations, are central to financial engineering and risk management. They allow financial engineer to price complex financial instruments. They allow risk managers to build the distribution of portfolios that are too complex to model analytically.

Simulation methods are quite flexible and are becoming easier to implement with technological advances in computing. Their drawbacks should not be underestimated, however. For all their elegance, simulation results depend heavily on the model's assumptions: the shape of the distribution, the parameters, and the pricing functions. Risk managers need to be keenly aware of the effect that errors in these assumptions can have on the results.

This chapter shows how Monte Carlo methods can be used for risk management. Section 4.1 introduces a simple case with just one source of risk. Section 4.2 shows how to apply these methods to construct value at risk (VAR) measures, as well as to price derivatives. Multiple sources of risk are then considered in Section 4.3.

4.1 SIMULATIONS WITH ONE RANDOM VARIABLE

Simulations involve creating artificial random variables with properties similar to those of the risk factors in the portfolio. These include stock prices, exchange rates, bond yields or prices, and commodity prices.

4.1.1 Simulating Markov Processes

In efficient markets, financial prices should display a random walk pattern. More precisely, prices are assumed to follow a **Markov process**, which is a particular stochastic process independent of its past history; the entire distribution of the future price relies on the current price only. The past history is irrelevant. These processes are built from the following components, described in order of increasing complexity.

- **The Wiener process.** This describes a variable Δz, whose change is measured over the interval Δt such that its mean change is zero and variance proportional to Δt:

$$\Delta z \sim N(0, \Delta t) \tag{4.1}$$

  If ϵ is a standard normal variable $N(0, 1)$, this can be written as $\Delta z = \epsilon\sqrt{\Delta t}$. In addition, the increments Δz are independent across time.
- **The generalized Wiener process.** This describes a variable Δx built up from a Wiener process, with in addition a constant trend a per unit time and volatility b:

$$\Delta x = a\Delta t + b\Delta z \tag{4.2}$$

  A particular case is the **martingale**, which is a zero drift stochastic process, $a = 0$, which leads to $E(\Delta x) = 0$. This has the convenient property that the expectation of a future value is the current value

$$E(x_T) = x_0 \tag{4.3}$$

- **The Ito process.** This describes a generalized Wiener process, whose trend and volatility depend on the *current* value of the underlying variable and time:

$$\Delta x = a(x, t)\Delta t + b(x, t)\Delta z \tag{4.4}$$

  This is a Markov process because the distribution depends only on the current value of the random variable x, as well as time. In addition, the innovation in this process has a normal distribution.

4.1.2 The Geometric Brownian Motion

A particular example of Ito process is the **geometric Brownian motion** (GBM), which is described for the variable S as

$$\Delta S = \mu S\Delta t + \sigma S\Delta z \tag{4.5}$$

The process is geometric because the trend and volatility terms are proportional to the current value of S. This is typically the case for stock prices, for which *rates of returns* appear to be more stationary than raw dollar returns, ΔS. It is also used for currencies. Because $\Delta S/S$ represents the capital appreciation only, abstracting from dividend payments, μ represents the expected total rate of return on the asset minus the rate of income payment, or dividend yield in the case of stocks.

Example: A Stock Price Process

Consider a stock that pays no dividends, has an expected return of 10% per annum, and volatility of 20% per annum. If the current price is \$100, what is the process for the change in the stock price over the next week? What if the current price is \$10?

The process for the stock price is

$$\Delta S = S(\mu \Delta t + \sigma \sqrt{\Delta t} \times \epsilon)$$

where ϵ is a random drawn from a standard normal distribution. If the interval is one week, or $\Delta t = 1/52 = 0.01923$, the mean is $\mu \Delta t = 0.10 \times 0.01923 = 0.001923$ and $\sigma \sqrt{\Delta t} = 0.20 \times \sqrt{0.01923} = 0.027735$. The process is $\Delta S = \$100(0.001923 + 0.027735 \times \epsilon)$. With an initial stock price at \$100, this gives $\Delta S = 0.1923 + 2.7735\epsilon$. With an initial stock price at \$10, this gives $\Delta S = 0.01923 + 0.27735\epsilon$. The trend and volatility are scaled down by a factor of 10.

This model is particularly important because it is the underlying process for the Black–Scholes formula. The key feature of this distribution is the fact that the volatility is proportional to S. This ensures that the stock price will stay positive. Indeed, as the stock price falls, its variance decreases, which makes it unlikely to experience a large downmove that would push the price into negative values. As the limit of this model is a normal distribution for $dS/S = d\ln(S)$, S follows a **lognormal distribution.**

This process implies that, over an interval $T - t = \tau$, the logarithm of the ending price is distributed as

$$\ln(S_T) = \ln(S_t) + (\mu - \sigma^2/2)\tau + \sigma\sqrt{\tau}\,\epsilon \tag{4.6}$$

where ϵ is a standardized normal variable.

Example: A Stock Price Process (Continued)

Assume the price in one week is given by $S = \$100\exp(R)$, where R has annual expected value of 10% and volatility of 20%. Construct a two-tailed 95% confidence interval for S.

The standard normal deviates that corresponds to a 95% confidence interval are $\alpha_{MIN} = -1.96$ and $\alpha_{MAX} = 1.96$. In other words, we have 2.5% in each tail. The 95% confidence band for R is then $R_{MIN} = \mu\Delta t - 1.96\sigma\sqrt{\Delta t} = 0.001923 - 1.96 \times 0.027735 = -0.0524$ and $R_{MAX} = \mu\Delta t + 1.96\sigma\sqrt{\Delta t} = 0.001923 + 1.96 \times 0.027735 = 0.0563$. This gives $S_{MIN} = \$100\exp(-0.0524) = \$94.89$, and $S_{MAX} = \$100\exp(0.0563) = \$105.79$.

Whether a lognormal distribution is much better than the normal distribution depends on the horizon considered. If the horizon is one day only, the choice of the lognormal versus normal assumption does not really matter. It is highly unlikely that the stock price would drop below zero in one day, given typical volatilities. On the other hand, if the horizon is measured in years, the two assumptions do lead to different results. The lognormal distribution is more realistic as it prevents prices form turning negative.

In simulations, this process is approximated by small steps with a normal distribution with mean and variance given by

$$\frac{\Delta S}{S} \sim N(\mu \Delta t, \sigma^2 \Delta t) \tag{4.7}$$

To simulate the future price path for S, we start from the current price S_t and generate a sequence of independent standard normal variables ϵ, for $i = 1, 2, \ldots, n$. The next price S_{t+1} is built as $S_{t+1} = S_t + S_t(\mu \Delta t + \sigma \epsilon_1 \sqrt{\Delta t})$. The following price S_{t+2} is taken as $S_{t+1} + S_{t+1}(\mu \Delta t + \sigma \epsilon_2 \sqrt{\Delta t})$, and so on until we reach the target horizon, at which point the price $S_{t+n} = S_T$ should have a distribution close to the lognormal.

Table 4.1 illustrates a simulation of a process with a drift (μ) of 0% and volatility (σ) of 20% over the total interval, which is divided into 100 steps.

The initial price is \$100. The local expected return is $\mu \Delta t = 0.0/100 = 0.0$ and the volatility is $0.20 \times \sqrt{1/100} = 0.02$. The second column shows the realization of a uniform $U(0, 1)$ variable. The value for the first step is $u_1 = 0.0430$. The next column transforms this variable into a normal variable with mean 0.0 and volatility of 0.02, which gives −0.0343. The price increment is then obtained by multiplying the random variable by the previous price, which gives −\$3.433. This generates a new value of $S_1 = \$100 - \$3.43 = \$96.57$. The process is repeated until the final price of \$125.31 is reached at the 100th step.

This experiment can be repeated as often as needed. Define K as the number of **replications**, or random trials. Figure 4.1 displays the first three trials. Each leads

TABLE 4.1 Simulating a Price Path

	Random Variable		Price	
Step i	Uniform u_i	Normal $\mu\Delta t + \sigma \Delta z$	Increment ΔS_i	Price S_{t+i}
0				100.00
1	0.0430	−0.0343	−3.433	96.57
2	0.8338	0.0194	1.872	98.44
3	0.6522	0.0078	0.771	99.21
4	0.9219	0.0284	2.813	102.02
...				
99				124.95
100	0.5563	0.0028	0.354	125.31

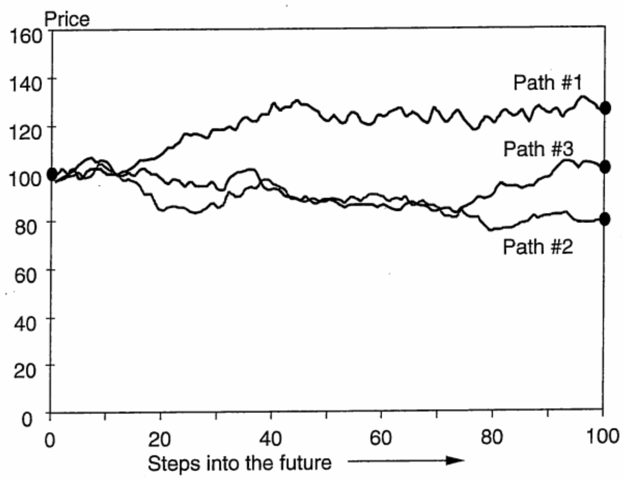

FIGURE 4.1 Simulating Price Paths

to a simulated final value S_T^k. This generates a distribution of simulated prices S_T. With just one step $n = 1$, the distribution must be normal. As the number of steps n grows large, the distribution tends to a lognormal distribution.

While very useful to model stock prices, this model has shortcomings. Price increments are assumed to have a normal distribution. In practice, we observe that price changes for most financial assets typically have fatter tails than the normal distribution. Returns may also experience changing variances.

In addition, as the time interval Δt shrinks, the volatility shrinks as well. This implies that large discontinuities cannot occur over short intervals. In reality, some assets experience discrete jumps, such as commodities, or securities issued by firms that go bankrupt. In such cases, the stochastic process should be changed to accommodate these observations.

EXAMPLE 4.1: FRM EXAM 2003—QUESTION 40

In the geometric Brownian motion process for a variable S,

I. S is normally distributed.
II. $d\ln(S)$ is normally distributed.
III. dS/S is normally distributed.
IV. S is lognormally distributed.

a. I only
b. II, III, and IV
c. IV only
d. III and IV

EXAMPLE 4.2: FRM EXAM 2002—QUESTION 126

Consider that a stock price S that follows a geometric Brownian motion $dS = aSdt + bSdz$, with b strictly positive. Which of the following statements is *false*?

a. If the drift a is positive, the price one year from now will be above today's price.
b. The instantaneous rate of return on the stock follows a normal distribution.
c. The stock price S follows a lognormal distribution.
d. This model does not impose mean reversion.

4.1.3 Drawing Random Variables

Most spreadsheets or statistical packages have functions that can generate uniform or standard normal random variables. This can be easily extended to distributions that better reflect the data, e.g., with fatter tails or non-zero skewness.

The methodology involves the inverse cumulative probability distribution function (p.d.f.). Take the normal distribution as an example. By definition, the cumulative p.d.f. $N(x)$ is always between 0 and 1. Because we have an analytical formula for this function, it can be easily inverted.

First, we generate a uniform random variable u drawn from $U(0, 1)$. Next, we compute x, such that $u = N(x)$, or $x = N^{-1}(u)$. For example, set $u = 0.0430$, as in the first line of Table 4.1. This gives $x = -1.717$.[1] Because u is less than 0.5, we verify that x is negative. The variable can be transformed into any normal variable by multiplying by the standard deviation and adding the mean. More generally, any distribution function can be generated as long as the cumulative distribution function can be inverted.

4.1.4 Simulating Yields

The GBM process is widely used for stock prices and currencies. Fixed-income products are another matter, however.

Bond prices display long-term reversion to the face value, which represents the repayment of principal at maturity (assuming there is no default). Such process is inconsistent with the GBM process, which displays no such mean reversion. The volatility of bond prices also changes in a predictable fashion, as duration shrinks to zero. Similarly, commodities often display mean reversion.

[1] In Excel, a uniform random variable can be generated with the function $u_i = \text{RAND}()$. From this, a standard normal random variable can be computed with $\text{NORMSINV}(u_i)$.

These features can be taken into account by modeling bond yields directly in a first step. In the next step, bond prices are constructed from the value of yields and a pricing function. The dynamics of interest rates r_t can be modeled by

$$\Delta r_t = \kappa(\theta - r_t)\Delta t + \sigma r_t{}^{\gamma} \Delta z_t \tag{4.8}$$

where Δz_t is the usual Wiener process. Here, we assume that $0 \leq \kappa < 1, \theta \geq 0$, $\sigma \geq 0$. Because there is only one stochastic variable for yields, the model is called a **one-factor model**.

This Markov process has a number of interesting features. First, it displays mean reversion to a long-run value of θ. The parameter κ governs the speed of mean reversion. When the current interest rate is high, i.e., $r_t > \theta$, the model creates a negative drift $\kappa(\theta - r_t)$ toward θ. Conversely, low current rates create a positive drift toward θ.

The second feature is the volatility process. This model includes the **Vasicek model** when $\gamma = 0$. Changes in yields are normally distributed because Δr is then a linear function of Δz, which is itself normal. The Vasicek model is particularly convenient because it leads to closed-form solutions for many fixed-income products. The problem, however, is that it could potentially lead to negative interest rates when the initial rate starts from a low value. This is because the volatility of the change in rates does not depend on the level, unlike that in the geometric Brownian motion.

Equation (4.8) is more general, however, because it includes a power of the yield in the variance function. With $\gamma = 1$, this is the **lognormal model**. Ignoring the trend, this gives $\Delta r_t = \sigma r_t \Delta z_t$, or $\Delta r_t / r_t = \sigma \Delta z_t$. This implies that the *rate of change* in the yield dr/r has a fixed variance. Thus, as with the GBM model, smaller yields lead to smaller movements, which makes it unlikely the yield will drop below zero. This model is more appropriate than the normal model when the initial yield is close to zero.

With $\gamma = 0.5$, this is the **Cox, Ingersoll, and Ross (CIR) model**. Ultimately, the choice of the exponent γ is an empirical issue. Recent research has shown that $\gamma = 0.5$ provides a good fit to the data.

This class of models is known as **equilibrium models**. They start with some assumptions about economic variables and imply a process for the short-term interest rate r. These models generate a predicted term structure, whose shape depends on the model parameters and the initial short rate. The problem with these models, however, is that they are not flexible enough to provide a good fit to today's term structure. This can be viewed as unsatisfactory, especially by practitioners who argue they cannot rely on a model that cannot be trusted to price today's bonds.

In contrast, **no-arbitrage models** are designed to be consistent with today's term structure. In this class of models, the term structure is an input into the parameter estimation. The earliest model of this type was the **Ho and Lee model:**

$$\Delta r_t = \theta(t)\Delta t + \sigma \Delta z_t \tag{4.9}$$

where $\theta(t)$ is a function of time chosen so that the model fits the initial term structure. This was extended to incorporate mean reversion in the **Hull and White model:**

$$\Delta r_t = [\theta(t) - ar_t]\Delta t + \sigma \Delta z_t \tag{4.10}$$

Finally, the **Heath, Jarrow, and Morton model** goes one step further and assumes that the volatility is a function of time.

The downside of these no-arbitrage models is that they do not impose any consistency between parameters estimated over different dates. The function $\theta(t)$ could be totally different from one day to the next, which is illogical. No-arbitrage models are also more sensitive to outliers, or data errors in bond prices used to fit the term structure.

4.1.5 Binomial Trees

Simulations are very useful to mimic the uncertainty in risk factors, especially with numerous risk factors. In some situations, however, it is also useful to describe the uncertainty in prices with discrete trees. When the price can take one of two steps, the tree is said to be **binomial.**

The binomial model can be viewed as a discrete equivalent to the geometric Brownian motion. As before, we subdivide the horizon T into n intervals $\Delta t = T/n$. At each "node," the price is assumed to go either up with probability p, or down with probability $1 - p$.

The parameters u, d, p are chosen so that, for a small time interval, the expected return and variance equal those of the continuous process. One could choose

$$u = e^{\sigma\sqrt{\Delta t}}, \quad d = (1/u), \quad p = \frac{e^{\mu\Delta t} - d}{u - d} \tag{4.11}$$

This matches the mean, for example,

$$E\left[\frac{S_1}{S_0}\right] = pu + (1 - p)d = \frac{e^{\mu\Delta t} - d}{u - d}u + \frac{u - e^{\mu\Delta t}}{u - d}d$$
$$= \frac{e^{\mu\Delta t}(u - d) - du + ud}{u - d} = e^{\mu\Delta t}$$

Table 4.2 shows how a binomial tree is constructed. As the number of steps increases, the discrete distribution of S_T converges to the lognormal distribution. This model will be used in a later chapter to price options.

TABLE 4.2 Binomial Tree

Step			
0	1	2	3
			u^3S
		u^2S	
	uS		u^2dS
S		udS	
	dS		d^2uS
		d^2S	
			d^3S

EXAMPLE 4.3: FRM EXAM 1999—QUESTION 25

The Vasicek model defines a risk-neutral process for r which is $dr = a(b - r)dt + \sigma dz$, where a, b, and σ are constant, and r represents the rate of interest. From this equation we can conclude that the model is a

a. Monte Carlo-type model
b. Single-factor term-structure model
c. Two-factor term-structure model
d. Decision tree model

EXAMPLE 4.4: FRM EXAM 1999—QUESTION 26

The term $a(b - r)$ in the previous question represents which term?

a. Gamma
b. Stochastic
c. Reversion
d. Vega

EXAMPLE 4.5: FRM EXAM 2000—QUESTION 118

Which group of term-structure models do the Ho-Lee, Hull-White, and Heath, Jarrow, and Morton models belong to?

a. No-arbitrage models
b. Two-factor models
c. Lognormal models
d. Deterministic models

EXAMPLE 4.6: FRM EXAM 2000—QUESTION 119

A plausible stochastic process for the short-term rate is often considered to be one where the rate is pulled back to some long-run average level. Which one of the following term-structure models does *not* include this characteristic?

a. The Vasicek model
b. The Ho-Lee model
c. The Hull-White model
d. The Cox-Ingersoll-Ross model

4.2 IMPLEMENTING SIMULATIONS

4.2.1 Simulation for VAR

Implementing Monte Carlo (MC) methods for risk management follows these steps:

1. Choose a stochastic process for the risk factor price S (i.e., its distribution and parameters, starting from the current value S_t).
2. Generate pseudo-random variables representing the risk factor at the target horizon, S_T.
3. Calculate the value of the portfolio at the horizon, $F_T(S_T)$.
4. Repeat steps 2 and 3 as many times as necessary. Call K the number of replications.

These steps create a distribution of values, $F_T^1, \ldots, F_T^K$, which can be sorted to derive the VAR. We measure the cth quantile $Q(F_T, c)$ and the average value

$\text{Ave}(F_T)$. If VAR is defined as the deviation from the expected value on the target date, we have

$$\text{VAR}(c) = \text{Ave}(F_T) - Q(F_T, c) \tag{4.12}$$

4.2.2 Simulation for Derivatives

Readers familiar with derivatives pricing will have recognized that this method is similar to the Monte Carlo method for valuing derivatives. In that case, we simply focus on the expected value on the target date discounted into the present:

$$F_t = e^{-r(T-t)}\,\text{Ave}(F_T) \tag{4.13}$$

Thus, derivatives valuation focuses on the discounted center of the distribution, while VAR focuses on the quantile on the target date.

Monte Carlo simulations have been long used to price derivatives. As will be seen in a later chapter, pricing derivatives can be done by assuming that the underlying asset grows at the risk-free rate r (assuming no income payment). For instance, pricing an option on a stock with expected return of 20% is done assuming that (1) the stock grows at the risk-free rate of 10% and (2) we discount at the same risk-free rate. This is called the **risk-neutral approach.**

In contrast, risk measurement deals with actual distributions, sometimes called **physical distributions.** For measuring VAR, the risk manager must simulate asset growth using the actual expected return μ of 20%. Therefore, risk management uses physical distributions, whereas pricing methods use risk-neutral distributions.

It should be noted that simulation methods are not applicable to all types of options. These methods assume that the value of the derivative instrument at expiration can be priced solely as a function of the end-of-period price S_T, and perhaps of its sample path. This is the case, for instance, with an Asian option, where the payoff is a function of the price *averaged* over the sample path. Such an option is said to be **path-dependent.**

Simulation methods, however, are inadequate to price American options, because such options can be exercised early. The optimal exercise decision, however, is complex to model because it should take into account *future* values of the option. This cannot be done with regular simulation methods, which only consider present and past information. Instead, valuing American options requires a **backward recursion**, for example with binomial trees. This method examines whether the option should be exercised or not, starting from the end and working backward in time until the starting time.

4.2.3 Accuracy

Finally, we should mention the effect of **sampling variability**. Unless K is extremely large, the empirical distribution of S_T will only be an approximation of the true distribution. There will be some natural variation in statistics measured from Monte Carlo simulations. Since Monte Carlo simulations involve *independent* draws, one can show that the standard error of statistics is inversely related to the square root of K. Thus more simulations will increase precision, but at a slow rate. For example, accuracy is increased by a factor of ten going from $K = 10$ to $K = 1{,}000$, but then requires going from $K = 1{,}000$ to $K = 100{,}000$ for the same factor of 10.

This accuracy issue is worse for risk management than for pricing, because the quantiles are estimated less precisely than the average. For VAR measures, the precision is also a function of the selected confidence level. Higher confidence levels generate fewer observations in the left tail and hence less-precise VAR measures. A 99% VAR using 1,000 replications should be expected to have only 10 observations in the left tail, which is not a large number. The VAR estimate is derived from the tenth and eleventh sorted number. In contrast, a 95% VAR is measured from the fiftieth and fifty-first sorted numbers, which is more precise. In addition, the precision of the estimated quantile depends on the shape of the distribution. Relative to a symmetric distribution, a short option position has negative skewness, or a long left tail. The observations in the left tail therefore will be more dispersed, making is more difficult to estimate VAR precisely.

Various methods are available to speed up convergence:

- **Antithetic Variable Technique.** This technique uses twice the same sequence of random draws from t to T. It takes the original sequence and changes the sign of all their values. This creates twice the number of points in the final distribution of F_T without running twice the number of simulations.
- **Control Variate Technique.** This technique is used to price options with trees when a similar option has an analytical solution. Say that f_E is a European option with an analytical solution. Going through the tree yields the values of an American and European option, F_A and F_E. We then assume that the error in F_A is the same as that in F_E, which is known. The adjusted value is $F_A - (F_E - f_E)$.
- **Quasi-Random Sequences.** These techniques, also called Quasi Monte Carlo (QMC), create draws that are not independent but instead are designed to fill the sample space more uniformly. Simulations have shown that QMC methods converge faster than Monte Carlo. In other words, for a fixed number of replications K, QMC values will be on average closer to the true value.

The advantage of traditional MC, however, is that it also provides a standard error, which is on the order of $1/\sqrt{K}$ because draws are independent. So, we have an idea of how far the estimate might be from the true value, which is useful to decide on the number of replications. In contrast, QMC methods give no measure of precision.

EXAMPLE 4.7: FRM EXAM 2005—QUESTION 67

Which of the following statements about Monte Carlo simulation is *false*?

a. Monte Carlo simulation can be used with a lognormal distribution.
b. Monte Carlo simulation can generate distributions for portfolios that contain only linear positions.
c. One drawback of Monte Carlo simulation is that it is computationally very intensive.
d. Assuming the underlying process is normal, the standard error resulting from Monte Carlo simulation is inversely related to the square root of the number of trials.

EXAMPLE 4.8: FRM EXAM 2007—QUESTION 66

A risk manager has been requested to provide some indication of accuracy of a Monte Carlo simulation. Using 1,000 replications of a normally distributed variable *S*, the relative error in the one-day 99% VAR is 5%. Under these conditions,

a. Using 1,000 replications of a long option position on *S* should create a larger relative error.
b. Using 10,000 replications should create a larger relative error.
c. Using another set of 1,000 replications will create an exact measure of 5.0% for relative error.
d. Using 1,000 replications of a short option position on *S* should create a larger relative error.

EXAMPLE 4.9: SAMPLING VARIATION

The measurement error in VAR, due to sampling variation, should be greater with

a. More observations and a high confidence level (e.g., 99%)
b. Fewer observations and a high confidence level
c. More observations and a low confidence level (e.g., 95%)
d. Fewer observations and a low confidence level

4.3 MULTIPLE SOURCES OF RISK

We now turn to the more general case of simulations with many sources of financial risk. Define N as the number of risk factors. If the factors S_j are independent, the randomization can be performed independently for each variable. For the GBM model,

$$\Delta S_{j,t} = S_{j,t-1}\mu_j \Delta t + S_{j,t-1}\sigma_j \epsilon_{j,t}\sqrt{\Delta t} \tag{4.14}$$

where the standard normal variables ϵ are independent across time and factor $j = 1, \ldots, N$.

In general, however, risk factors are correlated. The simulation can be adapted by, first, drawing a set of independent variables η, and, second, transforming them into correlated variables ϵ. As an example, with two factors only, we write

$$\begin{aligned} \epsilon_1 &= \eta_1 \\ \epsilon_2 &= \rho\eta_1 + (1-\rho^2)^{1/2}\eta_2 \end{aligned} \tag{4.15}$$

Here, ρ is the correlation coefficient between the variables ϵ. Because the η's have unit variance and are uncorrelated, we verify that the variance of ϵ_2 is one, as required

$$\mathrm{V}(\epsilon_2) = \rho^2\mathrm{V}(\eta_1) + [(1-\rho^2)^{1/2}]^2\mathrm{V}(\eta_2) = \rho^2 + (1-\rho^2) = 1$$

Furthermore, the correlation between ϵ_1 and ϵ_2 is given by

$$\mathrm{Cov}(\epsilon_1, \epsilon_2) = \mathrm{Cov}(\eta_1, \rho\eta_1 + (1-\rho^2)^{1/2}\eta_2) = \rho\mathrm{Cov}(\eta_1, \eta_1) = \rho$$

Defining ϵ as the *vector* of values, we verified that the covariance matrix of ϵ is

$$V(\epsilon) = \begin{bmatrix} \sigma^2(\epsilon_1) & \mathrm{Cov}(\epsilon_1, \epsilon_2) \\ \mathrm{Cov}(\epsilon_1, \epsilon_2) & \sigma^2(\epsilon_2) \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = R$$

Note that this covariance matrix, which is the expectation of squared deviations from the mean, can also be written as

$$V(\epsilon) = E[(\epsilon - E(\epsilon)) \times (\epsilon - E(\epsilon))'] = E(\epsilon \times \epsilon')$$

because the expectation of ϵ is 0. To generalize this approach to many more risk factors, however, we need a systematic way to derive the transformation in Equation (4.15).