CHAPTER THIRTY-ONE

Modern Portfolio Theory

Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality.

Milton Friedman, *"The Methodology of Positive Economics"*

This chapter starts with the **mean-variance theory** of portfolio selection. This theory provides a tractable framework for quantifying the risk–return trade-off of assets. We then investigate the equilibrium structure of asset prices. The result is the celebrated **Capital Asset Pricing Model** (**CAPM**, pronounced cap-m). The CAPM is the foundational quantitative model for measuring the risk of a security. Alternative asset pricing models based on factor analysis are also presented. The practically important concept of value at risk (VaR) for risk management concludes the chapter.

31.1 Mean–Variance Analysis of Risk and Return

Risk is the chance that expected returns will not be realized. We adopt standard deviation of the rate of return as the measure of risk.[1] This choice, although not without its critics, is standard in portfolio analysis and has nice statistical properties. Investors are presumed to prefer higher expected returns and lower variances.

Assume that there are n assets with random rates of return, $r_1, r_2, \ldots, r_n$. The expected values of these returns are $\overline{r}_i \equiv E[r_i]$. If we form a portfolio of these n assets by using (capitalization) weights $\omega_1, \omega_2, \ldots, \omega_n$, the portfolio's rate of return is

$$\boldsymbol{r} = \omega_1 r_1 + \omega_2 r_2 + \cdots + \omega_n r_n$$

with mean $\overline{\boldsymbol{r}} = \sum_{i=1}^{n} \omega_i \overline{r}_i$ and variance

$$\boldsymbol{\sigma}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i \omega_j \sigma_{ij} = \sum_{i \neq j} \omega_i \omega_j \sigma_{ij} + \sum_{i=1}^{n} \omega_i^2 \sigma_i^2,$$

where σ_i^2 represents the variance of r_i and σ_{ij} represents the covariance between r_i and r_j. Note that $\sigma_{ii} = \sigma_i^2$.

The portfolio's total risk as measured by its variance consists of (1) $\sum_{i \neq j} \omega_i \omega_j \sigma_{ij}$, the **systematic risk** associated with the correlations between the returns on the assets in the portfolio, and (2) $\sum_{i=1}^{n} \omega_i^2 \sigma_i^2$, the **specific** or **unsystematic risk** associated

https://doi.org/10.1017/CBO9780511546839.032 Published online by Cambridge University Press

with the individual variances alone. Every possible weighting scheme $\omega_1, \omega_2, \ldots, \omega_n$ with $\sum_{i=1}^{n} \omega_i = 1$ corresponds to a portfolio, with negative weights meaning short sales. The constraints $\omega_i \geq 0$ can be added to exclude short sales. A portfolio $\boldsymbol{\omega} \equiv [\,\omega_1, \omega_2, \ldots, \omega_n\,]^{\mathrm{T}}$ that satisfies all the specified constraints is said to be a **feasible portfolio**.

Interestingly, if the returns[2] of the assets are uncorrelated, i.e., $\sigma_{ij} = 0$ for $i \neq j$, the variance of the portfolio's return decreases toward zero as n increases, provided that the portfolio is well diversified. For example, with $\omega_i = 1/n$,

$$\boldsymbol{\sigma}^2 = \sum_{i=1}^{n} \omega_i^2 \sigma_i^2 = \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2} \leq \frac{\sigma_{\max}^2}{n},$$

where $\sigma_{\max} \equiv \max_i \sigma_i$. This shows the power of diversification. Diversification, however, has its limits when asset returns are correlated. To see this point, assume that (1) all the returns have the same variance s^2, (2) the return correlation is a constant z, hence $\sigma_{ij} = zs^2$ for $i \neq j$, and (3) $\omega_i = 1/n$. The variance of $\boldsymbol{r}$ then is

$$\boldsymbol{\sigma}^2 = \sum_{i \neq j} \frac{zs^2}{n^2} + \sum_{i=1}^{n} \frac{s^2}{n^2} = n(n-1)\,\frac{zs^2}{n^2} + \frac{s^2}{n} = zs^2 + (1-z)\,\frac{s^2}{n},$$

which cannot be reduced below the average covariance zs^2.

These two examples demonstrate that specific risk and systematic risk behave very differently as the number of assets included in the portfolio grows. In general, as the portfolio gets larger and is well diversified, the specific risk tends to zero, whereas the systematic risk converges to the average of all the covariances for all pairs of assets in the portfolio. Markowitz called this phenomenon the **law of the average covariance** [644]. Systematic risk therefore does *not* disappear with diversification.

Consider a two-dimensional diagram with the horizontal axis denoting standard deviation and the vertical axis denoting mean. This is called the **mean–standard deviation diagram**. Every feasible portfolio with mean return rate $\overline{\boldsymbol{r}}$ and standard deviation $\boldsymbol{\sigma}$ can be represented as a point at $(\boldsymbol{\sigma}, \overline{\boldsymbol{r}})$ on the diagram; it is an **obtainable mean–standard deviation combination**. The set of feasible points form the **feasible set**. In general, the feasible set is a solid two-dimensional region and convex to the left. Thus the straight line segment connecting any two points in the set does not cross the left boundary of the set. For a given expected rate of return, the feasible point with the smallest variance is the corresponding left boundary point. The left boundary of the feasible set is hence called the **minimum-variance set**, and the point on this set having the minimum variance is the **minimum-variance point** (**MVP**). Most investors will choose the portfolio with the smallest variance for a given mean. Such investors are risk averse because they seek to minimize risk as measured by the standard deviation. Similarly, most investors will choose the portfolio with the highest mean for a given level of standard deviation (i.e., the highest point on a given vertical line). Therefore only the subset of the minimum-variance set above the MVP will be of interest. An obtainable mean–standard deviation combination is **efficient** if no other obtainable combinations have either higher mean and no higher variance or less variance and no less mean. The set of efficient combinations is termed the **efficient frontier**, and the corresponding portfolios are termed the **efficient portfolios**. See Fig. 31.1.

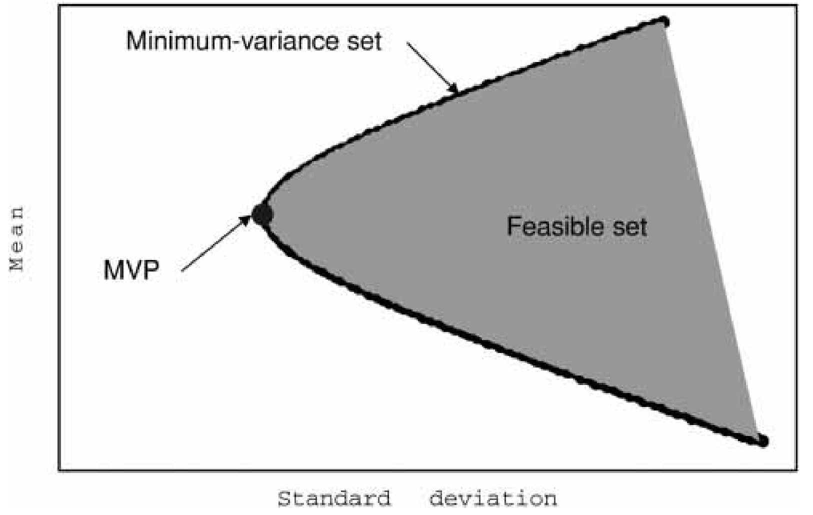

Figure 31.1: Feasible, minimum-variance, and efficient sets. The points in the minimum-variance set (that are above the MVP) form the efficient frontier, which is also called the **efficient set**. When short sales are not allowed, the feasible set is bounded because its mean lies within $[\min_i \overline{r}_i, \max_i \overline{r}_i]$ and its standard deviation lies within $[0, \max_i \sigma_i]$ (see Exercise 31.1.3).

Here is the mathematical formulation for the minimum-variance portfolio with a given mean value $\boldsymbol{\overline{r}}$ that is due to Markowitz in 1952 [641]:

$$\begin{aligned} &\text{minimize} \quad (1/2)\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_i\omega_j\sigma_{ij}, \\ &\text{subject to} \quad \sum_{i=1}^{n}\omega_i\overline{r}_i = \boldsymbol{\overline{r}}, \\ &\qquad\qquad\quad \sum_{i=1}^{n}\omega_i = 1. \end{aligned}$$

Short selling can be prohibited if $\omega_i \geq 0$ for $i = 1, 2, \ldots, n$, is added to the constraints. (The factor $1/2$ in front of the variance will simplify the analysis later.) The preceding **Markowitz problem** is a quadratic programming problem. It is a single-period investment theory that specifies the trade-off between the mean and the variance of a portfolio's rate of return.[3]

The Markowitz problem can be solved as follows. The weights ω_i and the two **Lagrange multipliers** λ and μ for an efficient portfolio satisfy

$$\begin{aligned} &\sum_{j=1}^{n}\sigma_{ij}\omega_j - \lambda\overline{r}_i - \mu = 0 \quad \text{for } i = 1, 2, \ldots, n, \\ &\sum_{i=1}^{n}\omega_i\overline{r}_i = \boldsymbol{\overline{r}}, \\ &\sum_{i=1}^{n}\omega_i = 1. \end{aligned}$$

There are $n+2$ equations with $n+2$ unknowns: $\omega_1, \omega_2, \ldots, \omega_n, \lambda, \mu$. Because the equations are linear, they can be easily solved (see Fig. 19.2). If the goal is to obtain the highest return for a given level of variance σ_{p}^2, then the problem becomes

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} \omega_i \bar{r}_i, \\ \text{subject to} \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i \omega_j \sigma_{ij} = \sigma_{\mathrm{p}}^2, \\ & \sum_{i=1}^{n} \omega_i = 1. \end{aligned}$$

Sophisticated quadratic programming techniques are needed to solve it.

Striking conclusions can be drawn from the mean–variance framework. Suppose that two solutions are available: (1) $(\boldsymbol{\omega}_1, \lambda_1, \mu_1)$ with expected return rate $\bar{\boldsymbol{r}}_1$ and (2) $(\boldsymbol{\omega}_2, \lambda_2, \mu_2)$ with expected return rate $\bar{\boldsymbol{r}}_2$. Direct substitution shows that $(\alpha\boldsymbol{\omega}_1 + (1-\alpha)\,\boldsymbol{\omega}_2, \alpha\lambda_1 + (1-\alpha)\,\lambda_2, \alpha\mu_1 + (1-\alpha)\,\mu_2)$ is also a solution to the $n+2$ equations and corresponds to the expected return rate $\alpha\bar{\boldsymbol{r}}_1 + (1-\alpha)\,\bar{\boldsymbol{r}}_2$. Thus the combined portfolio $\alpha\boldsymbol{\omega}_1 + (1-\alpha)\,\boldsymbol{\omega}_2$ also represents a point in the minimum-variance set. To use this result, suppose that $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are two different portfolios in the minimum-variance set. Then as α varies over $-\infty < \alpha < \infty$, the portfolios defined by $\alpha\boldsymbol{\omega}_1 + (1-\alpha)\,\boldsymbol{\omega}_2$ sweep out the entire minimum-variance set. In particular, if $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are efficient, they will generate all other efficient points. This is the **two-fund theorem**. Hence all investors seeking efficient portfolios need consider investing in combinations of only these two funds instead of individual stocks. This conclusion rests on the assumptions, among others, that everyone cares about only means and variances, that everyone has the same assessment of the parameters (means, variances, and covariances), that short selling is allowed, and that a single-period framework is appropriate.

➤ **Exercise 31.1.1** Express the efficient portfolio in matrix form.

➤ **Exercise 31.1.2** Construct a portfolio with zero risk from two perfectly negatively correlated assets without short sales.

➤ **Exercise 31.1.3** Let $C \equiv [\,\sigma_{ij}\,]$ be a positive definite matrix. (1) Prove that $\max_i \sigma_{ii}$ is the maximum value of $\sum_i \sum_j \omega_i \omega_j \sigma_{ij}$ under the constraints $\sum_i \omega_i = 1$ and $\omega_i \geq 0$. (2) How about the minimum value under the same constraints? (You may assume that the row sums of C^{-1} are all nonnegative.)

➤ **Exercise 31.1.4** Let $P(t)$ denote the asset price at time t. Define $r(T) \equiv [\,P(T)/P(0)\,] - 1$ as the holding period rate of return for a period of length T and $r_{\mathrm{c}}(T) \equiv \ln(P(T)/P(0))$ as the continuous holding period rate of return for the same period. Under the assumption that asset prices are lognormally distributed, derive the relations between the mean and the variance of $r(T)$ and those of $r_{\mathrm{c}}(T)$.

➤ **Exercise 31.1.5** Consider a portfolio P of n assets each following an independent geometric Brownian motion process with identical mean and variance, $dS_i/S_i = \mu dt + \sigma\, dW_i$. Each asset has the same weight of $1/n$ in the portfolio. Show that this portfolio's expected rate of return, $E[\,\ln(P(t)/P(0))\,]/t$, exceeds each individual asset's expected rate of return, $E[\ln(S_i(t)/S_i(0))]/t$, by $(1-1/n)\sigma^2/2$. (Volatility is thus not synonymous with risk.)

31.1.1 Adding the Riskless Asset

The riskless asset by definition has a return that is certain; its return has zero volatility. The riskless return's covariance with any risky asset's return is thus zero. The presence of the riskless asset in a portfolio implies lending or borrowing cash at the riskless rate: Lending means a long position in the asset, whereas borrowing means a short position. Clearly the riskless asset has to be a zero-coupon bond whose maturity matches the investment horizon.

The shape of the feasible set changes dramatically when the riskless asset is available. Let r_f denote the riskless rate of return. Start with the feasible set defined by *risky* assets. Now for each portfolio in this set, say portfolio A, form combinations with the riskless asset. These new combinations trace out the infinite straight line originating at the riskless point, passing through the risky portfolio, and continuing indefinitely: the r_f–A ray in Fig. 31.2. There is a ray of this type for every portfolio in the feasible set. The totality of these rays forms a triangularly shaped feasible set. If borrowing of the riskless asset is not allowed, we can adjoin only the line segment between the riskless asset and points in the original feasible set but cannot extend the line further. The inclusion of these line segments leads to a feasible set with a straight-line front edge but a rounded top: the r_f–P–Q curve in Fig. 31.2. Note that investors who hold some riskless assets invest the remaining funds in portfolio P, as they are on the r_f–P segment.

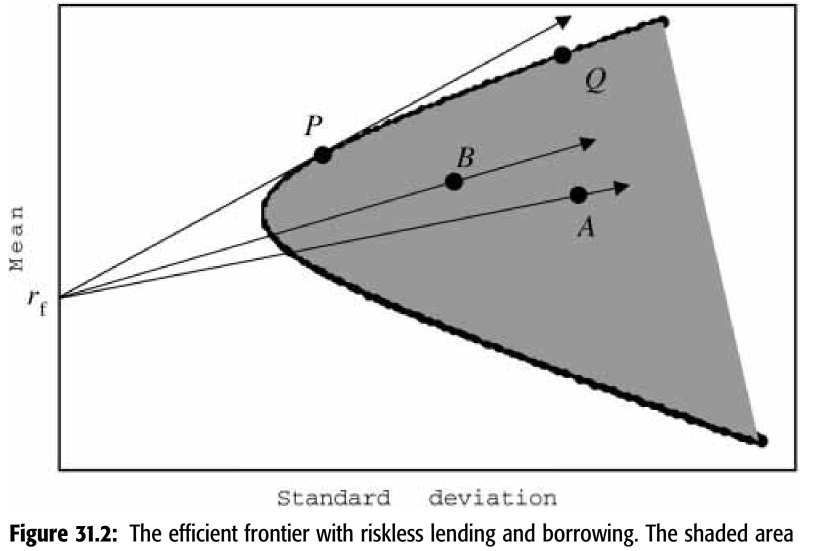

Figure 31.2: The efficient frontier with riskless lending and borrowing. The shaded area is the feasible set defined by risky assets. The line segment between r_f and A consists of combinations of portfolio A and lending, whereas the line segment beyond A consists of combinations of portfolio A and borrowing. The equation for the line is $y = r_f + x(\overline{r}_A - r_f)/\sigma_A$. The same observation can be made of any risky portfolio such as B, P, and Q. The ray through the tangent portfolio P defines the efficient frontier.

A special portfolio, denoted by P in Fig. 31.2, lies on the tangent point between the feasible set and a ray passing through r_{f}. When both borrowing and lending of the riskless asset are available, the efficient frontier is precisely this ray. Any efficient portfolio therefore can be expressed as a combination of P and the riskless asset. We have thus proved Tobin's **one-fund theorem**, which says there is a single fund of risky assets such that every efficient portfolio can be constructed as a combination of the fund and the riskless asset.

Identifying the tangent point P is computationally easy. For any point $(\sigma, \overline{\boldsymbol{r}})$ in the feasible set defined by risky assets, we can draw a line between the riskless asset and that point as in Fig. 31.2. The slope is equal to $\theta \equiv (\overline{\boldsymbol{r}} - r_{\mathrm{f}})/\sigma$, which has the interesting interpretation of the excess return per unit of risk. The tangent portfolio is the feasible point that maximizes θ. Assign weights $\omega_1, \omega_2, \ldots, \omega_n$ to the n risky assets such that $\sum_{i=1}^{n} \omega_i = 1$. The weight on the riskless asset in the tangent fund is zero. As a result, $\overline{\boldsymbol{r}} - r_{\mathrm{f}} = \sum_{i=1}^{n} \omega_i (\overline{r}_i - r_{\mathrm{f}})$, and

$$\theta = \frac{\sum_{i=1}^{n} \omega_i (\overline{r}_i - r_{\mathrm{f}})}{\sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \omega_i \omega_j}}.$$

Setting the derivative of θ with respect to each ω_j equal to zero leads to the equations

$$\lambda \sum_{i=1}^{n} \sigma_{ij} \omega_i = \overline{r}_j - r_{\mathrm{f}}, \quad j = 1, 2, \ldots, n,$$

where $\lambda \equiv \sum_{i=1}^{n} \omega_i (\overline{r}_i - r_{\mathrm{f}}) / (\sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \omega_i \omega_j) = (\overline{\boldsymbol{r}} - r_{\mathrm{f}})/\sigma^2$. Making the substitution $\nu_i = \lambda \omega_i$ for each i simplifies the preceding equations to

$$\sum_{i=1}^{n} \sigma_{ij} \nu_i = \overline{r}_j - r_{\mathrm{f}}, \quad j = 1, 2, \ldots, n. \tag{31.1}$$

We solve these linear equations for the ν_is (see Fig. 19.2) and determine ω_i by setting $\omega_i = \nu_i / (\sum_{j=1}^{n} \nu_j)$. A negative ω_i means that asset i needs to be sold short.

If riskless lending and borrowing are disallowed, the whole efficient frontier can be traced out by solving Eq. (31.1) for all possible riskless rates, because an efficient portfolio is a tangent portfolio to a ray extending from *some* riskless rate (consult Fig. 31.2 again). However, there is a better way. Observe that ν_i are linear in r_{f}; in other words, $\nu_i = c_i + d_i r_{\mathrm{f}}$ for some constants c_i and d_i. We can find c_i and d_i by first solving Eq. (31.1) for ν_i under two different r_{f}s, say r'_{f} and r''_{f}. The solutions ν'_i and ν''_i correspond to two efficient portfolios. Now we solve

$$\begin{aligned} \nu'_i &= c_i + d_i r'_{\mathrm{f}}, \\ \nu''_i &= c_i + d_i r''_{\mathrm{f}} \end{aligned}$$

for the unknown c_i and d_i for each i. By treating r_{f} as a variable and varying it, we can trace out the entire frontier. Just as the two-fund theorem says, two efficient portfolios suffice to determine the frontier.

➤ **Exercise 31.1.6** What would the one-fund theorem imply about trading volumes?

31.1.2 Alternative Efficient Portfolio Selection Models

In the **Black model**, portfolios are chosen subject only to $\sum_{i=1}^{n} \omega_i = 1$. In the **standard portfolio selection model**, short sales are disallowed, and the constraints are

$$\sum_{i=1}^{n} \omega_i = 1,$$
$$\omega_i \geq 0, \quad i = 1, 2, \ldots, n.$$

By law or by policy, there may be restrictions on the amounts that can be invested in any one security. To handle them, one may augment the standard model with upper bounds:

$$\sum_{i=1}^{n} \omega_i = 1,$$
$$\omega_i \geq 0, \quad i = 1, 2, \ldots, n,$$
$$\omega_i \leq u_i, \quad i = 1, 2, \ldots, n.$$

In the **Tobin–Sharpe–Lintner model**, the portfolios are chosen subject to

$$\sum_{i=1}^{n+1} \omega_i = 1,$$
$$\omega_i \geq 0, \quad i = 1, 2, \ldots, n.$$

The variable ω_{n+1} represents the amount lent (or borrowed if ω_{n+1} is negative). The covariances $\sigma_{n+1,i}$ are of course zero for $i = 1, 2, \ldots, n+1$. Limited borrowing can be modeled by the addition of the constraint $\omega_{n+1} \leq u_{n+1}$. In the **general portfolio selection model**, a portfolio is feasible if it satisfies

$$A\boldsymbol{\omega} = \boldsymbol{b},$$
$$\boldsymbol{\omega} \geq 0,$$

where A is any $m \times n$ matrix and $\boldsymbol{b}$ is an m-dimensional vector [642].

➤ **Exercise 31.1.7** Two portfolio selection models are **strictly equivalent** if they have the same set of obtainable mean–standard deviation combinations. Prove that any model that does not impose the nonnegative constraint on $\boldsymbol{\omega}$ is strictly equivalent to some general portfolio selection model, which does.

31.2 The Capital Asset Pricing Model

Imagine a world in which all investors are mean–variance portfolio optimizers and they share the same expectation as to expected returns, variances, and covariances. Also assume zero transactions cost. By the one-fund theorem, every investor will hold some amounts of the riskless asset and the same portfolio of risky assets. As all risky assets must be held by somebody, an immediate implication is that every investor holds the **market portfolio** in equilibrium regardless of one's degree of risk aversion. The market portfolio, which consists of all *risky* assets, is furthermore efficient (see Exercise 31.2.1).[4]

Given that the single efficient fund of risky assets is the market portfolio, the efficient frontier consists of a single straight line emanating from the riskless point

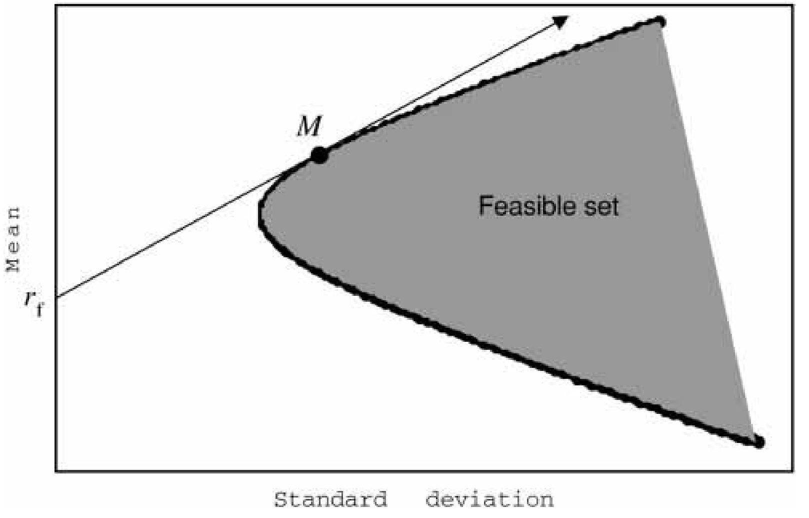

Figure 31.3: The capital market line. The point M stands for the market portfolio. The shaded set is the feasible set defined by risky assets. Investors can adjust the risk level by changing their holdings of riskless asset; for example, risk can be increased by holding negative amounts of the riskless asset.

and passing through the market portfolio. No complex computation is needed to determine the efficient frontier. This line is called the **capital market line**, which shows the relation between the expected rate of return and the risk for efficient portfolios (see Fig. 31.3). Prices should adjust so that efficient assets and portfolios fall on the line. Individual risky securities and inefficient portfolios, in contrast, will plot below the line. This is Sharpe's famous CAPM of 1964 [796],[5] which was independently arrived at by Lintner [605] and Mossin [680]. This model is fundamental to the equilibrium pricing of risky assets.

The capital market line states that

$$\overline{r} = r_{\mathrm{f}} + \frac{\overline{r}_M - r_{\mathrm{f}}}{\sigma_M}\, \sigma,$$

where $\overline{r}_M$ and σ_M are the expected value and the standard deviation of the market rate of return and $\overline{r}$ and σ are the expected value and the standard deviation of the rate of return of any efficient asset. Observe that as risk increases, the expected rate of return must also increase. The slope of the capital market line is $(\overline{r}_M - r_{\mathrm{f}})/\sigma_M$, which is called the **market price of risk**. It tells by how much the expected rate of return of an efficient portfolio must increase if the standard deviation of that rate increases by one unit. The market price of risk is also known as the **Sharpe ratio** [798].

The capital market line relates the expected rate of return of an efficient portfolio to its standard deviation, but it does not show how the expected rate of return of an individual asset relates to its individual risk. That relation is stated in the following theorem.

THEOREM 31.2.1 *If the market portfolio M is efficient, the expected return $\overline{r}_j$ of any asset j satisfies $\overline{r}_j - r_{\mathrm{f}} = \beta_j(\overline{r}_M - r_{\mathrm{f}})$, where $\beta_j \equiv \sigma_{j,M}/\sigma_M^2$ and $\sigma_{j,M} \equiv \mathrm{Cov}[\, r_j, r_M\,]$.*

The value β_i is referred to as the **beta** of an asset. An asset's beta is all that needs to be known about its risk characteristics. The value $\overline{r}_i - r_{\mathrm{f}}$ is the expected excess

Company	*Beta*	*Company*	*Beta*
America Online	2.43	Intel	1.03
AT&T	0.82	Merck	0.87
Citigroup	1.68	Microsoft	1.49
General Motors	1.01	Sun Micro.	1.19
IBM	1.07	Wal-Mart	1.20

Figure 31.4: Betas of some U.S. corporations. America Online merged with Time-Warner in 2001. Source: Standard & Poor's, May 8, 2000.

rate of return of asset i. It is the amount by which the rate of return is expected to exceed the riskless rate. Likewise, $\overline{r}_M - r_f$ is the expected excess rate of return of the market portfolio. The CAPM says that the expected excess rate of return of an asset is proportional to the expected excess rate of return of the market portfolio, and the proportionality factor is beta. Beta, not volatility, is the measure of a security's risk, and the method of beating the market is to assume greater risk, i.e., beta. Figure 31.4 shows the betas of some U.S. corporations. We can estimate beta by regressing the excess return on the asset against the excess return on the market.

The CAPM formula in Theorem 31.2.1 shows a linear relation between beta and the expected rate of return for all assets whether they are efficient or not. This relationship, when plotted on a beta expected-return diagram, is termed the **security market line**. All assets fall on the security market line; in particular, the market is the point at $\beta = 1$.

Essentially the same arguments go through even if there is no riskless asset (see Exercise 31.2.12). The role of the riskless rate of return is then played by the mean rate of return in which the line in Fig. 31.3 intercepts the axis of mean rate of return.

➤ **Exercise 31.2.1** Verify that the market portfolio is efficient.

➤ **Exercise 31.2.2** Prove the security market line formula in Theorem 31.2.1.

31.2.1 More on the CAPM

The portfolio beta is the weighted average of the betas of the individual assets in the portfolio. Specifically, suppose a portfolio contains n assets with the weights $\omega_1, \omega_2, \dots, \omega_n$. The rate of return of the portfolio is $r \equiv \sum_i \omega_i r_i$. Hence $\text{Cov}[\, r, r_M \,] = \sum_i \omega_i \sigma_{i,M}$. It follows immediately that the portfolio beta equals $\sum_i \omega_i \beta_i$.

Write asset i's rate of return as

$$r_i = r_f + \beta_i (r_M - r_f) + \epsilon_i, \tag{31.2}$$

where $E[\, \epsilon_i \,] = 0$ by the CAPM. Now take the covariance of r_i with r_M in Eq. (31.2) to yield

$$\sigma_{i,M} = \beta_i^2 \sigma_M^2 + \text{Cov}[\, \epsilon_i, r_M \,] = \sigma_{i,M} + \text{Cov}[\, \epsilon_i, r_M \,].$$

Therefore $\text{Cov}[\, \epsilon_i, r_M \,] = 0$ and

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \text{Var}[\, \epsilon_i \,]. \tag{31.3}$$

It is important to note that the total risk σ_i^2 is a sum of two parts. The first part, $\beta_i^2\sigma_M^2$, is the systematic risk. This is the risk associated with the market as a whole, also called the **market risk**. It cannot be reduced by diversification because every asset with nonzero beta contains this risk. The second part $\mathrm{Var}[\,\epsilon_i\,]$ is the specific risk. This risk is uncorrelated with the market and can be reduced by diversification. Only the systematic risk has any bearing on returns.

Consider an asset on the capital market line with a beta of β and an expected rate of return equal to $\overline{r} = r_{\mathrm{f}} + \beta(\overline{r}_M - r_{\mathrm{f}})$. This asset, which is efficient, must be equivalent to a combination of the market portfolio and the riskless asset. Its standard deviation is therefore $\beta\sigma_M$, which implies it has only systematic risk but no specific risk by Eq. (31.3). Now consider another asset with the same beta β. According to the CAPM, its expected rate of return must be $\overline{r}$. However, if it carries specific risk, it will not fall on the capital market line. The specific risk is thus the distance by which the portfolio lies below the capital market line.

Although stated in terms of expected returns, the CAPM is also a pricing model. Suppose an asset is purchased at a known price P and later sold at price Q. The rate of return is $r \equiv (Q - P)/P$. By the CAPM,

$$\frac{\overline{Q} - P}{P} = r_{\mathrm{f}} + \beta(\overline{r}_M - r_{\mathrm{f}}),$$

where β is the beta of the asset. Solve for P to obtain

$$P = \frac{\overline{Q}}{1 + r_{\mathrm{f}} + \beta(\overline{r}_M - r_{\mathrm{f}})}. \tag{31.4}$$

Hence the CAPM can be used to decide whether the price for a stock is "right." Note that the risk-adjusted interest rate is $r_{\mathrm{f}} + \beta(\overline{r}_M - r_{\mathrm{f}})$, not r_{f}.

Equation (31.4) can take another convenient form. The value of beta is

$$\beta = \frac{\mathrm{Cov}[\,r, r_M\,]}{\sigma_M^2} = \frac{\mathrm{Cov}[\,(Q/P) - 1, r_M\,]}{\sigma_M^2} = \frac{\mathrm{Cov}[\,Q, r_M\,]}{P\sigma_M^2}.$$

Substituting this into pricing formula (31.4) and dividing by P yields

$$1 = \frac{\overline{Q}}{P(1 + r_{\mathrm{f}}) + \mathrm{Cov}[\,Q, r_M\,]\,(\overline{r}_M - r_{\mathrm{f}})/\sigma_M^2}.$$

Solve for P again to obtain

$$P = \frac{1}{1 + r_{\mathrm{f}}}\left\{\overline{Q} - \frac{\mathrm{Cov}[\,Q, r_M\,]\,(\overline{r}_M - r_{\mathrm{f}})}{\sigma_M^2}\right\}. \tag{31.5}$$

This demonstrates it is the asset's covariance with the market that is relevant for pricing.

➤ **Exercise 31.2.3** For an asset uncorrelated with the market (that is, with zero beta), the CAPM says its expected rate of return is the riskless rate even if this asset is very risky with a large standard deviation. Why?

➤ **Exercise 31.2.4** If an asset has a negative beta, the CAPM says its expected rate of return should be less than the riskless rate even if this asset is very risky with a large standard deviation. Why? (For example, we saw in Chap. 24 that IO strips earn less than the riskless rate despite their high riskiness.)

https://doi.org/10.1017/CBO9780511546839.032 Published online by Cambridge University Press

➤ **Exercise 31.2.5** Why must all portfolios with the same expected rate of return but different total risks fall on the same point on the security market line?

➤ **Exercise 31.2.6** (1) Verify that pricing formula (31.4) is linear (the price of the sum of two assets is the sum of their prices, and the price of a multiple of an asset is the same multiple of the price). (2) Derive the same results from the no-arbitrage principle.

31.2.2 Portfolio Insurance

Portfolio insurance is a trading strategy that protects a portfolio from market declines but without losing the opportunity to participate in market rallies – in a word, a protective put [772]. Using puts to protect a portfolio from falling below a specified floor is a simple example of *static* portfolio insurance. Alternatives to static schemes are dynamic strategies that create synthetic options with stocks and bonds. Dynamic strategies, however, generate high transactions costs. This problem was mitigated by the introduction of stock index futures. Compared with the underlying assets, futures can be traded at much lower transactions costs in achieving the desired mixture of risky and riskless assets.[6]

Let the value of the index be S and each put be on \$100 times the index. Consider a diversified portfolio with a beta of β. If for each $100 \times S$ dollars in the portfolio, one put contract is purchased with strike price X, the value of the portfolio is protected against the possibility of the index's falling below the floor of X. Our goal is to implement this protective put. Specifically, to protect each dollar of the portfolio against falling below W at time T, we buy β put contracts for each $100 \times S$ dollars in the portfolio. Note that the total number of puts bought is $\beta V/(S \times 100)$, where V is the current value of the portfolio. The strike price X is the index value when the portfolio value reaches W.

Let r be the interest rate and q the dividend yield. Suppose that the index reaches S_T at time T. The excess return of the index over the riskless interest rate is $(S_T - S)/S + q - r$, and the excess return of the portfolio over the riskless interest rate is $\beta((S_T - S)/S + q - r)$. The return from the portfolio is therefore $\beta[\,(S_T - S)/S + q - r\,] + r$, and the increase in the portfolio value net of the dividends is $\beta[\,(S_T - S)/S + q - r\,] + r - q$. Therefore the portfolio value per dollar of the original value is

$$1 + \beta\left(\frac{S_T - S}{S} + q - r\right) + r - q = \beta\,\frac{S_T}{S} + (\beta - 1)(q - r - 1). \tag{31.6}$$

Choose X to be the S_T that makes Eq. (31.6) equal W; in other words,

$$X = [\,W + (q - r - 1)(1 - \beta)\,]\,\frac{S}{\beta}.$$

From Eq. (31.6), the portfolio value is less than W by $\beta(\Delta S/S)$ if and only if the index value is less than X by $\beta(\Delta S/S)(S/\beta) = \Delta S$. Exercising the options therefore induces a matching cash inflow of

$$\beta\,\frac{\Delta S}{100 \times S} \times 100 = \beta\,\frac{\Delta S}{S}.$$

The strategy's cost is $P\beta V/(S \times 100)$, where P is the put premium with strike price X.

Clearly a higher strike price provides a higher floor of WV dollars at a greater cost. This trade-off between the cost of insurance and the level of protection is typical of any insurance. The total wealth of course has a floor of

$$WV - \frac{P\beta V}{S \times 100}.$$

EXAMPLE 31.2.2 Start with $S = 1000$, $\beta = 1.5$, $q = 0.02$, and $r = 0.07$ for a period of 1 year. We have the following relations between the index value and the portfolio value per dollar of the original value.

Index value in a year	*1200*	*1100*	*1000*	*900*	*800*
Portfolio value in a year	1.275	1.125	0.975	0.825	0.675

For example, if the portfolio starts at \$1 million and the insured value is \$0.825 million, then $(1.5 \times 1{,}000{,}000)/(100 \times 1{,}000) = 15$ put contracts with a strike price of 900 should be purchased.

➤ **Exercise 31.2.7** Redo Example 31.2.2 with $S = 1000$, $\beta = 2$, $q = 0.01$, and $r = 0.05$.

➤ **Exercise 31.2.8** Consider a portfolio worth \$1,000 times the S&P 500 Index and with a beta of 1.0 against the index. Argue that buying 10 put index options with a strike price of 1,000 insures against the portfolio value's dropping below \$1,000,000.

➤ **Exercise 31.2.9** A mutual fund manager believes that the market is going to be relatively calm in the near future and writes a covered index call. Analyze it by following the same logic as that of the protective put.

➤ **Exercise 31.2.10** A bank offers the following financial product to a mutual fund manager planning to buy a certain stock in the near future. If the stock price is over \$50, the manager buys it at \$50. If the stock price is below \$40, the manager buys it at \$40. If the stock price is between the two, the manager buys at the spot price. Analyze the underlying options.

31.2.3 Critical Remarks

Fire those CAPM-peddling consultants.
—Louis Lowenstein [620]

Although the CAPM is widely used by practitioners [592], many of its assumptions have been controversial. It assumes either normally distributed asset returns or quadratic utility functions. It furthermore assumes that investors care about only the mean and the variance of returns, which implies that they view upside and downside risks with equal distaste. In reality, portfolio returns are not, strictly speaking, normally distributed, and investors seem to distinguish between upside and downside risks. The theory posits, unrealistically, that everyone has identical information about the returns of all assets and their covariances. Even if this assumption were valid, it would not be easy to obtain accurate data. Usually, the variances and covariances can be accurately estimated, but not the expected returns (see Example 20.1.1). Unfortunately, errors in means are more critical than errors in variances, and errors in variances are more critical than errors in covariances [204]. The assumption that all investors share a common investment horizon is rarely the case in practice.

The CAPM assumes that all assets can be bought and sold on the market. The assets include not just securities, but also real estate, cash, and even human capital. Because the market portfolio is difficult to define, in reality proxies for the market portfolio are used [799]. The trouble is that different proxies result in different beta estimates for the same security (see Exercise 31.2.12). Finally, a single risk factor does not seem adequate for describing the cross section of expected returns [145, 336, 424, 635, 636, 666].

➤ **Exercise 31.2.11** Why are security analysts' 1-year forecasts worse than 5-year ones?

➤ **Exercise 31.2.12** Prove that using any efficient portfolio for the risky assets as the proxy for the market portfolio results in linear relations between the expected rates of return and the betas, just as in the CAPM.

31.3 Factor Models

The mean–variance theory requires that many parameters be estimated: n for the expected returns of the assets and $n(n+1)/2$ for their covariances. Luckily, asset returns can often be explained by a much smaller number of underlying sources of randomness called factors. A factor model represents the connection between factors and individual returns. In this section a factor model of the return process for asset pricing is presented, the **Arbitrage Pricing Theory** (**APT**).

31.3.1 Single-Factor Models

We start with single-factor models. Suppose there are n assets with rates of return, $r_1, r_2, \dots, r_n$. There is a single factor f, which is a random quantity such as the return on a stock index for the holding period. The rates of return and the factor are related by

$$r_i = a_i + b_i f + \epsilon_i, \quad i = 1, 2, \dots, n,$$

where a_i and b_i are constants. The b_is are the factor loadings or **factor betas**, and they measure the sensitivity of the return to the factor. Without loss of generality, let $E[\,\epsilon_i\,] = 0$. Assume that ϵ_i are uncorrelated with f: $\mathrm{Cov}[\, f, \epsilon_i\,] = 0$. Furthermore, assume that they are uncorrelated with each other, i.e., $E[\,\epsilon_i \epsilon_j\,] = 0$ for $i \neq j$.[7] Any correlation between asset returns thus arises from a common response to the factor. The variance of ϵ_i is denoted by $\sigma^2_{\epsilon_i}$ and that of f by σ^2_f. There is a total of $3n+2$ parameters: $a_i, b_i, \sigma^2_{\epsilon_i}, \overline{f}$, and σ^2_f. The following results are straightforward:

$$\begin{aligned}
\overline{r}_i &= a_i + b_i \overline{f}, \\
\sigma_i^2 &= b_i^2 \sigma_f^2 + \sigma^2_{\epsilon_i}, \\
\mathrm{Cov}[\, r_i, r_j\,] &= b_i b_j \sigma_f^2, \quad i \neq j, \\
b_i &= \mathrm{Cov}[\, r_i, f\,] / \sigma_f^2.
\end{aligned}$$

The preceding simple covariance matrix leads to very efficient algorithms for the portfolio selection problems in Subsection 31.1.1 [317]. The single-factor model is due to Sharpe [795].

The return on a portfolio can be analyzed similarly. Consider a portfolio constructed with weights ω_i. Its rate of return is just

$$r = a + bf + \epsilon,$$

where $a \equiv \sum_{i=1}^{n} \omega_i a_i$, $b \equiv \sum_{i=1}^{n} \omega_i b_i$, and $\epsilon \equiv \sum_{i=1}^{n} \omega_i \epsilon_i$. The portfolio's beta b is hence the average of the underlying assets' betas b_i (recall the law of the average covariance). It is easy to verify that $E[\,\epsilon\,] = 0$, $\mathrm{Cov}[\,f, \epsilon\,] = 0$, and $\mathrm{Var}[\,\epsilon\,] = \sum_{i=1}^{n} \omega_i^2 \sigma_{\epsilon_i}^2$. Finally, the variance of r is

$$\sigma^2 = b^2 \sigma_f^2 + \mathrm{Var}[\,\epsilon\,],$$

similar to Eq. (31.3). Among the total risk above, the systematic part is $b^2\sigma_f^2$. The systematic risk, which is due to the $b_i f$ terms, results from the factor f that influences every asset and is therefore present even in a diversified portfolio. The $\mathrm{Var}[\,\epsilon\,]$ term represents the specific risk. The specific risk, which is due to the ϵ_i terms, can be made to go to zero through diversification. It is also called the **diversifiable risk**.

➤ **Exercise 31.3.1** Assume that the single factor f is the market rate of return, r_M. Write the return processes as $r_i - r_{\mathrm{f}} = \alpha_i + b_i(r_M - r_{\mathrm{f}}) + \epsilon_i$. As usual, $E[\,\epsilon_i\,] = 0$ and ϵ_i is uncorrelated with the market return. Show that $b_i = \mathrm{Cov}[\,r_i, r_M\,]/\mathrm{Var}[\,r_M\,]$, as in the CAPM.

31.3.2 Multifactor Models

Now there are two factors f_1 and f_2, and the rate of return of asset i takes the form

$$r_i = a_i + b_{i1} f_1 + b_{i2} f_2 + \epsilon_i.$$

As in Subsection 31.3.1, assume that $E[\,\epsilon_i\,] = 0$ and that ϵ_i is uncorrelated with the factors and ϵ_j for $j \neq i$. The formulas for the expected rates of return and covariances are

$$\begin{aligned}
\overline{r}_i &= a_i + b_{i1}\overline{f}_1 + b_{i2}\overline{f}_2,\\
\sigma_i^2 &= b_{i1}^2\sigma_{f_1}^2 + b_{i2}^2\sigma_{f_2}^2 + 2b_{i1}b_{i2}\,\mathrm{Cov}[\,f_1, f_2\,] + \sigma_{\epsilon_i}^2,\\
\mathrm{Cov}[\,r_i, r_j\,] &= b_{i1}b_{j1}\sigma_{f_1}^2 + b_{i2}b_{j2}\sigma_{f_2}^2 + (b_{i1}b_{i2} + b_{j1}b_{j2})\,\mathrm{Cov}[\,f_1, f_2\,], \quad i \neq j.
\end{aligned}$$

From the preceding equations,

$$\begin{aligned}
\mathrm{Cov}[\,r_i, f_1\,] &= b_{i1}\sigma_{f_1}^2 + b_{i2}\,\mathrm{Cov}[\,f_1, f_2\,],\\
\mathrm{Cov}[\,r_i, f_2\,] &= b_{i2}\sigma_{f_2}^2 + b_{i1}\,\mathrm{Cov}[\,f_1, f_2\,].
\end{aligned}$$

These give two equations that can be solved for b_{i1} and b_{i2}. Factor models with more than two factors are easy generalizations. For U.S. stocks, between 3 and 15 factors may be needed [623].

➤ **Exercise 31.3.2** Describe a procedure to convert a set of correlated factors into a set of uncorrelated factors, which are easier to handle.

31.3.3 The Arbitrage Pricing Theory (APT)

The factor-model framework leads to an alternative theory of asset pricing, Ross's APT, which is a theory about equilibrium under factor models [765]. The APT does not require that investors evaluate portfolios on the basis of means and variances. Neither is a quadratic utility function required. Instead, (1) the mean–variance framework is replaced with a factor model for returns, (2) investors are assumed to prefer a greater return to a lesser return when returns are certain, and (3) the universe of assets is assumed to be large.

Consider first a special case in which the rates of return observe the one-factor model:

$$r_i = a_i + b_i f.$$

This factor model has no residual errors, and the uncertainty associated with a return is due solely to the uncertainty in the factor f. Interestingly, the values of a_i and b_i must be related if arbitrage opportunities are to be excluded. Here is the argument. Consider two assets i and j with $b_i \neq b_j$. Now form a portfolio with weights $\omega_i \equiv \omega$ for asset i and $\omega_j \equiv 1 - \omega$ for asset j. Its rate of return is

$$r = \omega a_i + (1 - \omega)\, a_j + (\omega b_i + (1 - \omega)\, b_j)\, f.$$

If we select $\omega = b_j/(b_j - b_i)$ to make the coefficient of f zero, the rate of return r becomes

$$\lambda_0 \equiv \frac{a_i b_j - a_j b_i}{b_j - b_i}.$$

This portfolio is riskless because the equation for r contains no random elements. If there happens to be a riskless asset, then $\lambda_0 = r_{\mathrm{f}}$. Even if riskless assets do not exist, all portfolios constructed without dependence on f must have the same rate of return, λ_0. Now $\lambda_0(b_j - b_i) = a_i b_j - a_j b_i$, which can be rearranged as

$$\frac{a_j - \lambda_0}{b_j} = \frac{a_i - \lambda_0}{b_i}.$$

As this relation holds for all i and j, there is a constant c such that

$$\frac{a_i - \lambda_0}{b_i} = c$$

for all i.[8] The values of a_i and b_i are thus related by $a_i = \lambda_0 + b_i c$. The expected rate of return of asset i is now

$$\overline{r}_i = a_i + b_i \overline{f} = \lambda_0 + b_i c + b_i \overline{f} = \lambda_0 + b_i \lambda_1, \tag{31.7}$$

where $\lambda_1 \equiv c + \overline{f}$. We see that once the constants λ_0 and λ_1 are known, the expected return of an asset is determined entirely by the factor betas b_i. The above analysis can be generalized (see Exercise 31.3.4).

THEOREM 31.3.1 *Let there be n assets whose rates of return are governed by $m < n$ factors according to the equations $r_i = a_i + \sum_{j=1}^{m} b_{ij} f_j, i = 1, 2, \ldots, n$. Then there exist constants $\lambda_0, \lambda_1, \ldots, \lambda_m$ such that $\overline{r}_i = \lambda_0 + \sum_{j=1}^{m} b_{ij} \lambda_j$, $i = 1, 2, \ldots, n$. The value λ_i is called the market price of risk associated with factor f_i or simply the* **factor price**.

Next we consider general multifactor models with residual errors,

$$r_i = a_i + \sum_{j=1}^{m} b_{ij} f_j + \epsilon_i,$$

where $E[\epsilon_i] = 0$ and $\sigma_{\epsilon_i}^2 \equiv E[\epsilon_i^2]$. As before, ϵ_i is assumed to be uncorrelated with the factors and with the residual errors of other assets. Let us form a portfolio by using the weights $\omega_1, \omega_2, \ldots, \omega_n$ with $\sum_{i=1}^{n} \omega_i = 1$. The rate of return of the portfolio is

$$r = a + \sum_{j=1}^{m} b_j f_j + \epsilon,$$

where $a \equiv \sum_{i=1}^{n} \omega_i a_i$, $b_j \equiv \sum_{i=1}^{n} \omega_i b_{ij}$, and $\epsilon \equiv \sum_{i=1}^{n} \omega_i \epsilon_i$. Let $\sigma_{\epsilon_i} \leq S$ for some constant S for all i. Assume that the portfolio is *well diversified* in the sense that $\omega_i \leq W/n$ for some constant W for all i – no single asset dominates the portfolio. Then

$$\text{Var}[\epsilon] = \sum_{i=1}^{n} \omega_i^2 \sigma_{\epsilon_i}^2 \leq \frac{1}{n^2} \sum_{i=1}^{n} W^2 S^2 = \frac{W^2 S^2}{n} \to 0$$

as $n \to \infty$. Combined with the fact $E[\epsilon] = 0$, the residual error ϵ of a well-diversified portfolio selected from a very large number of assets is approximately zero.[9]

A riskless portfolio in terms of zero sensitivity to all factors was used in the proof of Theorem 31.3.1. We just showed that the portfolio remains riskless under the more general models as long as it is well diversified. The existence of a riskless well-diversified portfolio suffices to extend Theorem 31.3.1 to the more general models (see Exercise 31.3.5).

The APT and the CAPM are not directly comparable [289]. Neither of the two models' assumptions imply the other's. The CAPM makes strong assumptions about the probability distribution of assets' rates of return, agents' utility functions, or both. The APT on the other hand, makes strong assumptions about assets' equilibrium rates of return. However, because the APT does not identify the factors, the CAPM can be made consistent with the APT and vice versa. For example, consider a two-factor model $r_i = a_i + b_{i1} f_1 + b_{i2} f_2 + \epsilon_i$. Under the APT model, $\overline{r}_i = r_f + b_{i1}\lambda_1 + b_{i2}\lambda_2$. Let portfolio j $(j = 1, 2)$ have an expected return rate of $\lambda_j + r_f$ and a beta value of β_{f_j}. Clearly portfolio j's only source of risk is factor f_j. Because the CAPM says that $\lambda_j = \beta_{f_j}(\overline{r}_M - r_f)$,

$$\overline{r}_i = r_f + b_{i1}\beta_{f_1}(\overline{r}_M - r_f) + b_{i2}\beta_{f_2}(\overline{r}_M - r_f) = r_f + (b_{i1}\beta_{f_1} + b_{i2}\beta_{f_2})(\overline{r}_M - r_f).$$

The beta is thus a weighted sum of the underlying factors' betas with the factor betas as the weights. Different factor betas are the reason different assets have different betas.

➤ **Exercise 31.3.3** For the one-factor APT, what will become of λ_1 if the CAPM holds?

➤ **Exercise 31.3.4** Prove Theorem 31.3.1.

➤ **Exercise 31.3.5** Complete the proof of the APT under the general factor models.

31.4 Value at Risk

> Anyone that relied on so-called
> value-at-risk models has been crucified.
> —The Economist, 1999 [309]

Introduced in 1983, the VaR is an attempt to provide a single number for senior management that summarizes the total risk in a portfolio of financial assets. The VaR calculation is aimed at making a statement of the form: "We are c percent certain not to lose more than V dollars in the next m days." The variable V is the VaR of the portfolio. The VaR is therefore an estimate, with a given degree of confidence, of how much one can lose from one's portfolio over a given time horizon, or

$$\text{Prob}[\text{ change in portfolio value} \leq -\text{VaR}\,] = 1 - c,$$

where c is the confidence level (see Fig. 31.5). The VaR is usually calculated assuming "normal" market circumstances, meaning that extreme market conditions such as market crashes are not considered. It has become widely used by corporate treasurers and fund managers as well as financial institutions.

For the purposes of measuring the adequacy of bank capital, the Bank for International Settlements (BIS) sets the confidence level $c = 0.99$ and the time horizon $m = 10$ (days) [293]. Another interesting application is in investment evaluation. Here risk is viewed in terms of the impact of the prospective change on the overall value at risk, i.e., *incremental* VaR, and we go ahead with the investment if the incremental VaR is low enough relative to the expected return [283].

Suppose returns are normally distributed and independent on successive days.[10] We consider a single asset first. We assume that the stock price is S, whose daily volatility for the return rate $\Delta S/S$ is σ. Because the time horizon m is usually small, we assume that the expected price change is zero. The standard deviation of the stock price over this time horizon is $S\sigma\sqrt{m}$. The VaR of holding one unit of the stock is $2.326 \times S\sigma\sqrt{m}$ if the confidence level is 99% and $1.645 \times S\sigma\sqrt{m}$ if

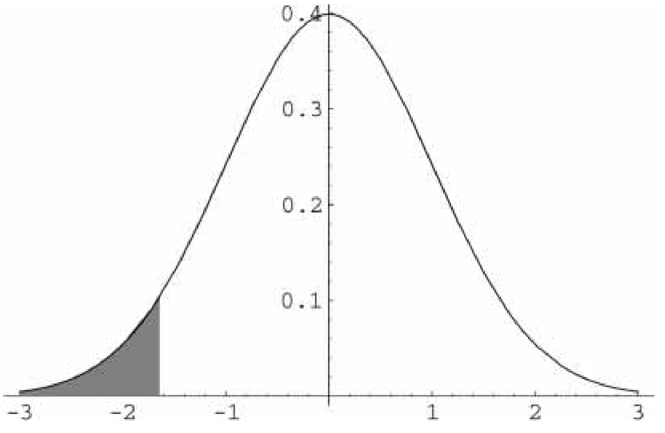

Figure 31.5: Confidence level and VaR. The diagram shows a confidence level of 95% under the standard normal distribution; the shaded area is 5% of the total area under the density function. It corresponds to 1.64485 standard deviations from the mean.

https://doi.org/10.1017/CBO9780511546839.032 Published online by Cambridge University Press

Confidence level (c)	*Number of standard deviations*
95%	1.64485
96%	1.75069
97%	1.88079
98%	2.05375
99%	2.32635

Figure 31.6: Confidence levels and standard deviations from the mean. The table samples confidence levels and their corresponding numbers of standard deviations from the mean when the random variable is normally distributed.

the confidence level is 95%. In general, the VaR is $-N^{-1}(1-c)$ times the standard deviation, or

$$-N^{-1}(1-c)\, S\sigma\sqrt{m},$$

where $N(\cdot)$ is the distribution function of the standard normal distribution (see Fig. 31.6). The preceding equation makes it easy to convert one horizon or confidence level to another. For example, the relation between 99% VaR and 95% VaR is

$$\text{VaR}\,(95\%) = \text{VaR}\,(99\%) \times (1.645/2.326).$$

Similarly, the variance of an m-day return should be m times the variance of a 1-day return. The m-day VaR thus equals $\sqrt{m}$ times the 1-day VaR, which is also called the **daily earnings at risk**. When m is not small, the expected annual rate of return μ needs to be considered. In that case, the drift $S\mu m/T$ is subtracted from the VaR when there are T trading days per annum.

Now consider a portfolio of assets. Assume that the changes in the values of asset prices have a multivariate normal distribution. Let the daily volatility of asset i be σ_i, let the correlation between the returns on assets i and j be ρ_{ij}, and let S_i be the market value of the positions in asset i. The VaR for the whole portfolio then is

$$-N^{-1}(1-c)\sqrt{m}\sqrt{\sum_i \sum_j S_i S_j \sigma_i \sigma_j \rho_{ij}}\,.$$

This way of computing the VaR is called the **variance–covariance approach** [518]. It was popularized by J.P. Morgan's RiskMetrics™ (1994).

The variance–covariance methodology may break down if there are derivatives in the portfolio because the returns of derivatives may not be normally distributed even if the underlying asset is. Nevertheless, if movements in the underlying asset are expected to be very small because, say, the time horizon is short, we may approximate the sensitivity of the derivative to changes in the underlying asset by the derivative's delta as follows. Consider a portfolio P of derivatives with a single underlying asset S. Recall that the delta of the portfolio, δ, measures the price sensitivity to S, or approximately $\Delta P/\Delta S$. The standard deviation of the distribution of the portfolio is $\delta S\sigma\sqrt{m}$, and its VaR is $-N^{-1}(1-c)\,\delta S\sigma\sqrt{m}$. In general when there are many underlying assets, the VaR of a portfolio containing options becomes

$$-N^{-1}(1-c)\sqrt{m}\sqrt{\sum_i \sum_j \delta_i \delta_j S_i S_j \sigma_i \sigma_j \rho_{ij}},$$

where δ_i denotes the delta of the portfolio with respect to asset i and S_i is the value of asset i. This is called the **delta approach** [527, 878]. The delta approach essentially treats a derivative as delta units of its underlying asset for the purpose of VaR calculation. This is not entirely unreasonable because such equivalence does hold instantaneously. It becomes questionable, however, as m increases.

Rather than using asset prices, VaR usually relies on a limited number of basic market variables that account for most of the changes in portfolio value [603]. As mentioned in Subsection 31.3.1, this greatly reduces the complexity related to the covariance matrix because only the covariances between the market variables are needed now. Typical market variables are yields or bond prices, exchange rates, and market returns. A basic instrument is then associated with each market variable. A security is now approximated by a portfolio of these basic instruments. Finally, its VaR is reduced to those of the basic instruments.

➤ **Exercise 31.4.1** What is the VaR of a futures contract on a stock?

➤ **Exercise 31.4.2** If the stock price follows $dS = S\mu\, dt + S\sigma\, dW$, what is its VaR τ years from now at c confidence?

31.4.1 Simulation

The Monte Carlo simulation is a general method to estimate the VaR, particularly for derivatives [571]. It works by computing the values of the portfolio over many sample paths, and the VaR is based on the distribution of the values. Figure 31.7 contains an algorithm for n asset prices following geometric Brownian motion:

$$\frac{dS_j}{S_j} = \mu_j\, dt + \sigma_j\, dW_j, \quad j = 1, 2, \dots, n,$$

where the n factors, dW_j, are correlated. As always, *actual* returns, not risk-neutral returns, should be used. For short time horizons, this distinction is not critical for most cases. In practice, to save computation time, a stock with a beta of β is mapped to a position in β times the index. Of course, this approach ignores the stock's specific risk. A related simulation method, called **historical simulation**, utilizes historical data [518]. It is identical to the Monte Carlo simulation except that the sample paths are generated by sampling the historical data as if they are to be repeated in the future.

Brute-force Monte Carlo simulation is inefficient when the number of factors is large. Fortunately, factor analysis and principal components analysis can often reduce the number of factors needed in the simulation [2, 509]. Let C denote the covariance matrix of the n factors $dW_1, dW_2, \dots, dW_n$. Let $u_i \equiv [\, u_{1i}, u_{2i}, \dots, u_{ni}\,]^{\mathrm{T}}$ be the eigenvectors of C and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the corresponding positive eigenvalues. Hence $\lambda_i u_i = C u_i$ for $i = 1, 2, \dots, n$. Recall that each eigenvalue indicates how much of the variation in the data its corresponding eigenvector explains. By the Schur decomposition theorem, the eigenvectors can be assumed to be orthogonal to each other. Normalize the eigenvectors such that $|u_i|^2 = \lambda_i$ and define

$$dZ_j \equiv \lambda_j^{-1} \sum_{k=1}^{n} u_{kj}\, dW_k.$$

It follows that

$$dW_i = \sum_{k=1}^{n} u_{ik}\, dZ_k,$$

VaR with Monte Carlo simulation:

```
input:    p̄, c, n, C[n][n], S[n], μ[n], Δt, m, N;
real      S[m+1][n], y[n], dW[n], P[n][n], p[N];
real      ξ();   // ξ() ~ N(0,1).
integer   i, j, k;
Let P be such that C = PPᵀ;   // See p. 248.
for (j = 0 to n−1) { S[0][j] := S[j]; }
for (k = 0 to N−1) {
          for (i = 1 to m) {
             for (j = 0 to n−1)
                y[j] := ξ() × √Δt;
             dW := Py;
             for (j = 0 to n−1) {
                S[i][j] := S[i−1][j] × ((1 + μ[j]) × Δt + √C[j][j] × dW[j]);
             }
          }
          Calculate the horizon portfolio value p[k];
}
Sort p[0], p[1], ..., p[N−1] in non-decreasing order;
return p̄ − p[⌊(1−c)N−1⌋];
```

Figure 31.7: VaR with Monte Carlo simulation. The expected rates of return $\mu[\,]$ and the covariances are annualized. There are n assets, the portfolio's initial value is $\overline{p}$, c is the confidence level, C is the covariance matrix for the annualized asset returns, m is the number of days until the horizon, the number of replications is N, and $S[\,]$ stores the initial asset prices. Recall that $C = PP^{\mathrm{T}}$ is the Cholesky decomposition of C. The portfolio's values at the horizon date are calculated and stored in $p[\,]$. Here we need pricing models and assume that early exercise is not possible during the period. The appropriate percentile is returned after sorting.

where $dZ_k\, dZ_j = 0$ for $j \neq k$ and $dZ_k\, dZ_k = dt$ (see Exercise 31.4.3, part (3)). If the empirical analysis shows that all but the first m principal components are small, then

$$dW_i \approx \sum_{k=1}^{m} u_{ik}\, dZ_k.$$

As a result, the asset price processes can be approximated by

$$\frac{dS_j}{S_j} \approx \mu_j\, dt + \sigma_j \sum_{k=1}^{m} u_{ik}\, dZ_k, \quad j = 1, 2, \ldots, n.$$

Only m orthogonal factors $dZ_1, dZ_2, \ldots, dZ_m$ remain.

➤ **Exercise 31.4.3** Prove that (1) $C = PP^{\mathrm{T}}$, where P's ith column is the eigenvector u_i, (2) $P^{-1} = \text{diag}[\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1}]\, P^{\mathrm{T}}$, (3) $P[dZ_1, dZ_2, \ldots, dZ_n]^{\mathrm{T}} = [dW_1, dW_2, \ldots, dW_n]^{\mathrm{T}}$, and (4) $P^{\mathrm{T}}P = \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$.

31.4.2 Critical Remarks

VaR relies on certain assumptions that are inconsistent with empirical evidence. Many implementations assume that asset returns are normally distributed. This simplifies the computation considerably but is inconsistent with the empirical evidence, which finds that many returns have fat tails, both left and right, at both daily and monthly time horizons. Extreme events are hence much more likely to occur in practice than would be predicted based on the assumption of normality [857]. A standard measure of tail fatness is kurtosis. Price jumps and stochastic volatility can be used to generate fat tails (see, e.g., Exercise 20.2.1) [293]. Although daily market returns are not normal [743], for longer periods, say 3 months, returns are quite close to being normally distributed [592].

The method of calculating the VaR depends on the horizon. A method yielding good results over a short horizon may not work well over longer horizons. The method of calculating the VaR also depends on asset types. If the portfolio contains derivatives, methods different from these used to analyze portfolios of stocks may be needed [464].

The ability to quantify risk exposure into a number represents the single most powerful advantage of the VaR. However, the VaR is extremely dependent on parameters, data, assumptions, and methodology. Although it should be part of an effective risk management program, the VaR is not sufficient to control risk [61, 464]. On occasion, it becomes necessary to quantify the magnitude of the losses that might accrue under events less likely than those analyzed in a standard VaR calculation. The procedures used to quantify potential loss exposures under such special circumstances are called **stress tests** [571]. A stress test measures the loss that could be experienced if a set of factors are exogenously specified.

31.4.3 VaR for Fixed-Income Securities

In contrast to stock prices, bond prices tend to move together because much of the movement is systematic, the common factor being the interest rate. For this reason, bond portfolio management does not require that the portfolios be well diversified. Instead, a few bonds of differing maturities can usually hedge the price fluctuations in any single bond or portfolio of bonds [91].

Duration (see Section 4.2) and key rate duration (see Section 27.5) were used to quantify the interest rate exposure of fixed-income portfolios and securities. A VaR methodology can also be based on duration. If S refers to the initial yield of a fixed-income instrument with duration D, the VaR for a long position in the instrument is $1.645 \times \sigma SD$ for a 95% confidence level. As before, VaR analysis requires parameters for the *actual* term structure dynamics. Simulation-based VaR usually conducts factor analysis before the actual simulation [804]. Three orthogonal factors seem to be sufficient (see Subsection 19.2.5).

The variance–covariance approach to the VaR is more complicated [470, 720]. First, an "equivalent" portfolio of standard zero-coupon bonds is obtained for each bond (this is called **cash flow mapping**). Then the historical volatility of spot rates and the correlations between them are used to construct a 95% confidence interval for the dollar return. It is difficult to apply this approach to securities with embedded options, however.

Additional Reading

In 1952, Markowitz and Roy independently published their papers that mark the era of modern portfolio theory [642, 644]. Our presentation of modern portfolio theory was drawn from [317, 623]. General treatment of mean–variance can be found in [643]. See [82, 174, 403, 760, 862] for additional information on beta and [81, 332, 399] for the issue of expected return and risk. The framework of modern portfolio theory can be applied to real estate [389, 407]. See [28] for modern portfolio theory's applicability in Japan. One of the reasons cited for the choice of standard deviation as the measure of risk is that it is easier to work with than the alternatives [799]. An interesting theory from experimental psychology, the **prospect theory**, says that an investor is much more sensitive to reductions in wealth than to increases, which is called **loss aversion** [533, 534]. Spreads can be used to profit from such behavioral "biases" [180]. Consult [592] for an approach beyond mean–variance analysis.

See [826] for development of the utility function. Optimization theory is discussed in [278, 687]. See [346, 470, 567, 646, 693] for additional information on portfolio insurance. Consult [317, 623, 673] for investment performance evaluation; there seems to be no consistent performance for mutual funds [634]. See [96, 360, 799] for security analysis, such as technical analysis and fundamental analysis, and [132] on market timing.

Consult [314] for the VaR of derivatives, [484, 691, 720, 771] for VaR when returns are not normally distributed, [8] for managing VaR using puts, [293, 390, 464, 484, 720, 804, 857] for the VaR of fixed-income securities. The **Cornish–Fisher expansion** is useful for correcting the skewness in distribution during VaR calculations [470, 522].

NOTES

1. Sometimes we use variance of return as the measure of risk for convenience.
2. "Rate of return" and "return" are often used interchangeably as only single-period analysis is involved.
3. Markowitz's Ph.D. dissertation was initially voted down by Friedman on the grounds that "It's not math, it's not economics, it's not even business administration." See [64, p. 60].
4. The S&P 500 Index often serves as the proxy for the market portfolio. **Index funds** are mutual funds that attempt to duplicate a stock market index. Offered in 1975 under the name of Vanguard Index Trust, the Vanguard 500 Index Fund is the first index mutual fund. It tracks the S&P 500 and became the largest mutual fund in April of 2000.
5. Sharpe sent the paper in 1962 to the *Journal of Finance*, but it was quickly rejected [64, p. 194].
6. Dynamic strategies rely on the market to supply the needed liquidity. The Crash of 1987 and the Russian and the LTCM crises of 1998 demonstrate that such liquidity may not be available at times of extreme market movements [308, 654]. As prices began to fall during the Crash of 1987, portfolio insurers sold stock index futures. This activity in the futures market led to more selling in the cash market as program traders attempted to arbitrage the spreads between the cash and futures markets. Further price declines led to more selling by portfolio insurers, and so on [567, 647].
7. The CAPM does not require that the residuals ϵ_i be uncorrelated; see Eq. (31.2).
8. See Eqs. (15.12) and (24.10) for similar arguments.
9. Chebyshev's inequality in Exercise 13.3.10, part (2), supplies the intuition.
10. "Return" means price change ΔS or simple rate of return $\Delta S/S$. This is consistent with the stochastic differential equation $\Delta S = S\mu\,\Delta t + S\sigma\sqrt{\Delta t}\,\xi$ when Δt is small.