CHAPTER 8

The Capital Asset Pricing Model

Chapter Outline

8.1 Introduction

The capital asset pricing model (CAPM) is an equilibrium theory built on the foundation of modern portfolio theory. It is, however, an equilibrium theory with a somewhat peculiar structure. This is true for a number of reasons:

1. First, the CAPM is a theory of financial equilibrium only. Investors take the various statistical quantities—means, variances, covariances—that characterize a security's return process as given. There is no attempt within the theory to link the return processes with events in the *real* side of the economy. In future model contexts, we shall generalize this feature (Chapter 11).

2. Second, as a theory of financial equilibrium, it makes the assumption that the supply of existing assets is equal to the demand for existing assets and, as such, that the currently observed asset prices are equilibrium ones. There is no attempt, however, to compute asset supply and demand functions explicitly. Only the equilibrium price vector is characterized. Let us elaborate on this point.

   Under the CAPM, portfolio theory informs us about the demand side. If individual i invests a fraction w_{ij} of his initial wealth Y_{0i} in asset j, the value of his asset j holding is $w_{ij}Y_{0i}$. Absent any information that he wishes to alter these holdings, we may interpret the quantity $w_{ij}Y_{0i}$ as his demand for asset j at the prevailing price vector. If there are I individuals in the economy, the total value of all holdings of asset j is $\sum_i^I w_{ij}Y_{0i}$; by the same remark we may interpret this quantity as aggregate demand. At equilibrium, one must have $\sum_i^I w_{ij}Y_{0i} = P_jQ_j$, where p_j is the prevailing equilibrium price per share of asset j, Q_j is the total number of shares outstanding and, consequently, p_jQ_j is the market capitalization of asset j. The CAPM derives its implications for prices by assuming that the actual economy-wide asset holdings are investors' aggregate optimal asset holdings.
3. Third, the CAPM expresses equilibrium in terms of relationships between the return distributions of individual assets and the return characteristics of the portfolio of all assets. We may view the CAPM as informing us, via modern portfolio theory, as to what asset return interrelationships must be in order for equilibrium asset prices to coincide with the observed asset prices.

   In what follows we first present an overview of the traditional approach to the CAPM. This is followed by a more general presentation that permits at once a more complete and more general characterization. In both presentations we depart from earlier notation and denote a security's mean return not by μ, but rather by $E(\tilde{r})$. This notation is more typical of the empirical literature that we review at the end of the chapter.

8.2 The Traditional Approach to the CAPM

To get useful results in this complex world of many assets, we have to make simplifying assumptions. The CAPM approach essentially hypothesizes (1) that all agents have the *same beliefs* about future returns (i.e., homogeneous expectations), and, in its simplest form, (2) that there is a risk-free asset, paying a safe return r_f at which investors can borrow or lend as much as they wish. These assumptions guarantee (Chapter 6) that the mean–variance efficient frontier is the same for every investor, and furthermore, by the separation theorem, that all investors' optimal portfolios have an identical structure: a fraction of initial wealth is invested in the risk-free asset, the rest in the tangency portfolio (two-fund separation). It is then possible to derive a few key characteristics of equilibrium asset and portfolio returns without detailing the underlying equilibrium structure, i.e., the demand for and supply of assets, or discussing their prices.

Because all investors acquire shares in the same risky tangency portfolio T and make no other risky investments, all existing risky assets must belong to T by the definition of an equilibrium. Indeed, if some asset k were not found in T, there would be no demand for it; yet, it is assumed to exist in positive supply. Supply would then exceed demand, which is inconsistent with assumed financial market equilibrium. The same reasoning implies that the share of any asset j in portfolio T must correspond to the ratio of the market value of that asset p_jQ_j to the market value of all assets $\sum_{j=1}^{J} p_jQ_j$. This, in turn, guarantees that tangency portfolio T must be nothing other than the market portfolio M, the portfolio of all existing assets where each asset appears in a proportion equal to the ratio of its market value to the total market capitalization.

This simple reasoning leads to a number of useful conclusions:

1. The market portfolio is efficient because it is on the efficient frontier.
2. All individual optimal portfolios are located on the line originating at point $(0, r_f)$ and going through $(E\tilde{r}_M, \sigma_M)$, which is also the locus of all efficient portfolios (Figure 8.1). This locus is usually called the **capital market line** or CML.
3. The slope of the CML is $(E\tilde{r}_M - r_f)/\sigma_M$. It tells us that an investor considering a marginally riskier efficient portfolio would obtain, in exchange, an increase in expected return of $(E\tilde{r}_M - r_f)/\sigma_M$. This is the price of, or reward for, risk taking—the price of risk as applicable to efficient portfolios. In other words, for efficient portfolios, we have the simple linear relationship in Eq. (8.1).

$$E\tilde{r}_p = r_f + \frac{E\tilde{r}_M - r_f}{\sigma_M}\sigma_p \tag{8.1}$$

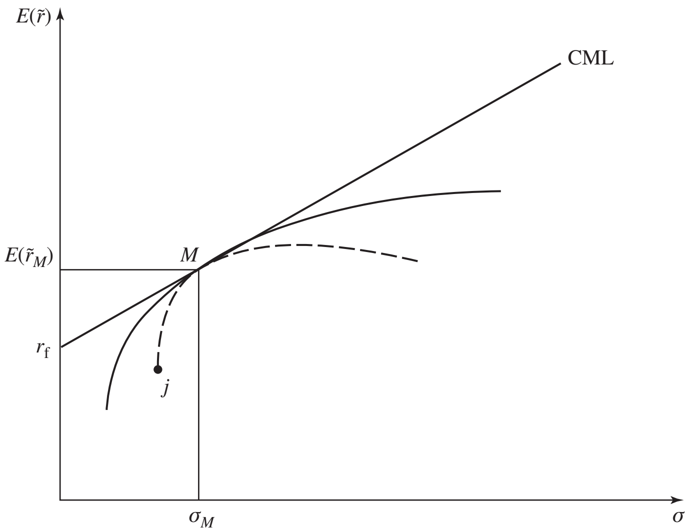

Figure 8.1
The CML.

The CML applies only to efficient portfolios. What can be said of an arbitrary asset j that does not belong to the efficient frontier? To discuss this essential part of the CAPM, we first rely on Eq. (8.2), formally derived in Appendix 8.1, and limit our discussion to its intuitive implications:

$$E\tilde{r}_j = r_f + (E\tilde{r}_M - r_f)\frac{\sigma_{jM}}{\sigma_M^2} \tag{8.2}$$

Let us define $\beta_j = \sigma_{jM}/\sigma_M^2$, i.e., the ratio of the covariance between the returns on asset j and the returns on the market portfolio divided by the variance of the market returns. We can thus rewrite Eq. (8.2) as Eq. (8.3).

$$E\tilde{r}_j = r_f + \left(\frac{E\tilde{r}_M - r_f}{\sigma_M}\right)\beta_j\sigma_M = r_f + \left(\frac{E\tilde{r}_M - r_f}{\sigma_M}\right)\rho_{jM}\sigma_j \tag{8.3}$$

Comparing Eqs. (8.1) and (8.3), we obtain one of the major lessons of the CAPM: only the fraction ρ_{jM} of the total risk of an asset j, σ_j, is remunerated by the market. The remaining fraction $(1 - \rho_{jM})\sigma_j$ is not: it is "diversified away" when asset j is placed in the market portfolio. By "diversified away," we mean that some of j's return variation is offset or canceled out by variation in the returns to other assets in the market portfolio.

To see this intuitively, let us first be reminded that under the CAPM, every investor holds only the market portfolio $(T = M)$, and thus the relevant risk for an investor can only be the market's standard deviation σ_M. As a consequence, what is important to the investor is the risk contribution of asset j to the risk of the market portfolio, i.e., the extent to which the inclusion of asset j into the overall portfolio M increases the latter's standard deviation. This marginal contribution of j to the overall portfolio risk σ_M is appropriately measured by $\rho_{jM}\sigma_j\ (= \beta_j\sigma_M)$ as the following equivalence demonstrates:

$$\begin{aligned} \sigma_M^2 &= \sum_{j=1}^{J} w_j \mathrm{cov}(\tilde{r}_j, \tilde{r}_M) = \sum_{j=1}^{J} w_j \rho_{jM}\sigma_j\sigma_M, \text{ and thus} \\ \sigma_M &= \sum_{j=1}^{J} w_j(\rho_{jM}\sigma_M) \end{aligned} \tag{8.4}$$

Accordingly, the risk premium on a given asset j is the market price of risk, $(E\tilde{r}_M - r_f)/\sigma_M$, multiplied by the measure of the quantity of j's relevant risk: $\rho_{jM}\sigma_j$ (or $\beta_j\sigma_M) \leq \sigma_j$. As such, $\rho_{jM}\sigma_j$ measures the **systematic risk** of asset j, systematic in the sense that it is the portion of j's risk contributing to variation in the market portfolio's return.[1] While the β_j of asset j

[1] Alternatively, and perhaps more informatively, systematic risk is also referred to as the "undiversifiable risk" or "market risk" of asset j.

is more typically referred to as asset j's systematic risk measure (we will also use this language), it more precisely measures ($\beta_j = \sigma_{jM}/\sigma_M^2 = (\rho_{jM}\sigma_j)/\sigma_M$) the systematic risk of asset *j relative* to the systematic risk of the overall market (loosely, the (value-weighted) average systematic risk of the assets in M).

A comparison of Eqs. (8.1) and (8.3) reaffirms that an efficient portfolio is one for which all diversifiable risks have been eliminated, i.e., where $\rho_{jM} = 1$. For an efficient portfolio, total risk and systematic risk are one and the same: within the available universe of assets, there are no further opportunities for diversification.

These ideas can also be developed by writing, without loss of generality, the excess return on asset j as a linear function of the excess return on the market portfolio plus a random term that is independent of the market return:

$$\tilde{r}_j - r_{\mathrm{f}} = \alpha_j + \beta_j(\tilde{r}_M - r_{\mathrm{f}}) + \tilde{\varepsilon}_j \tag{8.5}$$

If we go one step further and regard Eq. (8.5) as an OLS (ordinary least squares) regression equation, we obtain the standard regression estimate of the coefficient on the market, $\hat{\beta}$, where

$$\hat{\beta}_j = \frac{\hat{\sigma}_{jM}}{\hat{\sigma}_M^2}$$

an identification identical in form to our theoretical beta, thus accounting for the "beta label." Going on two steps further, the standard OLS variance decomposition yields

$$\hat{\sigma}_j^2 = \hat{\beta}_j^2 \hat{\sigma}_M^2 + \hat{\sigma}_{\varepsilon_j}^2 \tag{8.6}$$

Since $\hat{\beta}_j^2 \hat{\sigma}_M^2 = (\hat{\rho}_{jM}\hat{\sigma}_j)^2$ we may, accordingly, identify the first term in the variance decomposition as the systematic risk of j and identify the second as its diversifiable risk (relative to M). It is this latter term that disappears when j is placed in M.

Finally, Eq. (8.3) can equivalently be rewritten as

$$E\tilde{r}_j - r_f = (E\tilde{r}_M - r_{\mathrm{f}})\beta_j \tag{8.7}$$

which says that the expected excess return or the risk premium on an asset j is proportional to its β_j. Equation (8.7) defines the **security market line** or SML. It is depicted in Figure 8.2.

We conclude the overview of the CAPM by drawing attention to the fact that Eq. (8.5), when viewed as an OLS regression, may also be interpreted as a single factor model of security return generation. As such the excess return on the market portfolio constitutes the single explanatory factor with the "CAPM beta" serving as its factor sensitivity.

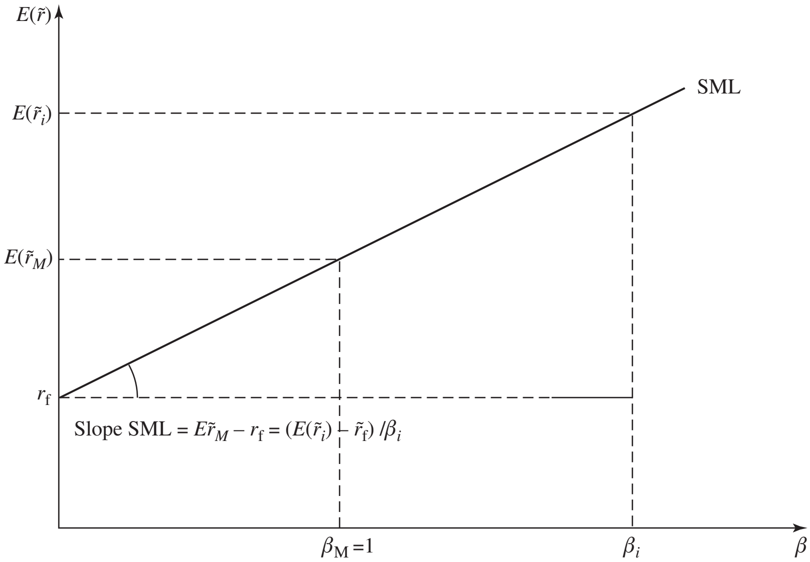

Figure 8.2
The SML.

8.3 Valuing Risky Cash Flows with the CAPM

We are now in position to make use of the CAPM not only to price assets but also to value nontraded risky cash flows such as those arising from an investment project. The traditional approach to this problem proposes to value an investment project at its present value price (i.e., at the appropriately discounted sum of the expected future cash flows). The logic is straightforward: to value a project equal to the present value of the expected future cash flows discounted at a particular rate is to price the project in a manner such that, at its present value price, it is expected to earn that discount rate. The appropriate rate, in turn, must be the analyst's estimate of the rate of return on *other* financial assets that represent title to cash flows *similar* in risk and timing to that of the project in question. This strategy has the consequence of pricing the project to pay the prevailing competitive rate for its risk class.

Enter the CAPM, which makes a definite statement regarding the appropriate discount factor to be used or, equivalently, the risk premium that should be applied to discount expected future cash flows. Strictly speaking, the CAPM is a one-period model; it is thus formally appropriate to use it only for one-period cash flows or projects.[2] In practice,

[2] Merton (1973) derives a multiperiod continuous-time version of the CAPM, provided cash flows follow a Markov process from period to period.

its use is more general and a multiperiod cash flow is typically viewed as the sum of one-period cash flows, each of which can be evaluated with the approach we now describe.

Consider some project j with cash-flow pattern

$$\begin{array}{l|l} t & t+1 \\ \hline -p_{j,t} & \tilde{C}F_{j,t+1} \end{array}$$

The link with the CAPM is immediate once we define the rate of return on project j. For a financial asset, we would naturally write $\tilde{r}_{j,t+1} = (\tilde{p}_{j,t+1} + \tilde{d}_{j,t+1} - p_{j,t})/p_{j,t}$, where $\tilde{d}_{j,t}$ is the dividend or any flow payment associated with the asset between date t and $t+1$. Similarly, if the initial value of the project with cash flow $\tilde{C}F_{j,t+1}$ is $p_{j,t}$, the return on the project is $\overline{r}_{j,t+1} = (\tilde{C}F_{j,t+1} - p_{j,t})/p_{j,t}$.

One thus has

$$1 + E(\tilde{r}_j) = E\left(\frac{\tilde{C}F_{j,t+1}}{p_{j,t}}\right) = \frac{E(\tilde{C}F_{j,t+1})}{p_{j,t}}, \text{ and by the CAPM,}$$

$$E(\tilde{r}_j) = r_{\mathrm{f}} + \beta_j(E(\tilde{r}_M) - r_{\mathrm{f}}), \text{ or}$$

$$1 + E(\tilde{r}_j) = 1 + r_{\mathrm{f}} + \beta_j(E(\tilde{r}_M) - r_{\mathrm{f}}), \text{ or}$$

$$\frac{E(\tilde{C}F_{j,t+1})}{p_{j,t}} = 1 + r_{\mathrm{f}} + \beta_j(E(\tilde{r}_M) - r_{\mathrm{f}}). \text{ Thus,}$$

$$p_{j,t} = \frac{E(\tilde{C}F_{j,t+1})}{1 + r_{\mathrm{f}} + \beta_j(E(\tilde{r}_M) - r_{\mathrm{f}})}$$

According to the CAPM, the project is thus priced at the present value of its expected cash flows discounted at the risk-adjusted rate appropriate to its risk class as identified by its β_j.

As discussed in Chapter 2, there is another potential approach to the pricing problem. It consists in altering the *numerator* of the pricing equations (the sum of expected cash flows) so that it is permissible to discount at the risk-free rate. This approach is based on the concept of certainty equivalent, which we discussed in Chapter 3. The idea is simple: If we replace each element of the future cash flow by its *CE*, it is clearly permissible to discount at the risk-free rate. Since we are interested in equilibrium valuations, however, we need a market certainty equivalent rather than an individual investor one. It turns out that this approach raises exactly the same set of issues as the more common one just considered: an equilibrium asset pricing model is required to tell us what market risk premium is appropriate to deduct from the expected cash flow to obtain its *CE*. Again the CAPM helps solve this problem.[3]

In the case of a one-period cash flow, transforming period-by-period cash flows into their market certainty equivalents can be accomplished in a straightforward fashion by applying the CAPM equation to the rate of return expected on the project. With $\tilde{r}_j = (\tilde{C}F_{j,t+1}/p_{j,t}) - 1$, the CAPM implies

$$E\left(\frac{\tilde{C}F_{j,t+1}}{p_{j,t}} - 1\right) = r_f + \beta_j(E(\tilde{r}_M) - r_f) = r_f + \frac{\operatorname{cov}\left(\frac{\tilde{C}F_{j,t+1}}{p_{j,t}} - 1, \tilde{r}_M\right)}{\sigma_M^2}(E(\tilde{r}_M) - r_f)$$

or

$$E\left(\frac{\tilde{C}F_{j,t+1}}{p_{j,t}} - 1\right) = r_f + \frac{1}{p_{j,t}}\operatorname{cov}(\tilde{C}F_j^{t+1}, \tilde{r}_M)\left[\frac{E(\tilde{r}_M) - r_f}{\sigma_M^2}\right]$$

Solving for $p_{j,t}$ yields

$$p_{j,t} = \frac{E(\tilde{C}F_{j,t+1}) - \operatorname{cov}(\tilde{C}F_{j,t+1}, \tilde{r}_M)\left[\frac{E(\tilde{r}_M) - r_f}{\sigma_M^2}\right]}{1 + r_f}$$

which one may also write

$$p_{j,t} = \frac{E(\tilde{C}F_{j,t+1}) - p_{j,t}\beta_j[E(\tilde{r}_M) - r_f]}{1 + r_f}$$

By appropriately transforming the expected cash flows, i.e., by subtracting what we have called an insurance premium (in Chapter 3), one can thus discount at the risk-free rate. The equilibrium certainty equivalent can thus be defined using the CAPM relationship. Note the information requirements in the procedure: if what we are valuing is indeed a one-off, nontraded cash flow, the estimation of the β_j, or of $\operatorname{cov}(\tilde{C}F_{j,t+1}, \tilde{r}_M)$, is far from straightforward; in particular, it cannot be based on historical data since there are none for the project at hand. It is here that the standard prescription calls for identifying a traded asset that can be viewed as similar in the sense of belonging to the same risk class. The estimated β for that traded asset is then to be used as an approximation in the above valuation formulas.

In the sections that follow, we first generalize the analysis of the efficient frontier presented in Chapter 6 to the $N \geq 2$ asset case. Such a generalization will require the use of elementary matrix algebra and is one of those rare situations in economic science where a

3 As does the arbitage pricing theory (APT); see Chapter 14.

more general approach yields a greater specificity of results. We will, for instance, be able to detail a version of the CAPM without a risk-free asset. This is then followed by the derivation of the standard CAPM where a risk-free asset is present.

As noted in the introduction, the CAPM is essentially an interpretation that we are able to apply to the efficient frontier. Not surprisingly, therefore, we begin this task with a return to characterizing that frontier.

8.4 The Mathematics of the Portfolio Frontier: Many Risky Assets and No Risk-Free Asset

Notation. Assume $N \geq 2$ risky assets; assume further that no asset has a return that can be expressed as a linear combination of the returns to a subset of the other assets (the returns are linearly independent). Let V denote the variance–covariance matrix, in other words, $V_{ij} = \text{cov}(r_i, r_j)$; by construction V is symmetric. Linear independence in the above sense implies that V^{-1} exists. Let w represent a column vector of portfolio weights for the N assets. The expression $w^T V w$ then represents the portfolio's return variance: $w^T V w$ is always positive (i.e., V is positive definite).

Let us illustrate this latter assertion in the two-asset case

$$
\begin{aligned}
w^T V w &= (\, w_1 \;\; w_2 \,) \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = (w_1 \sigma_1^2 + w_2 \sigma_{21} \quad w_1 \sigma_{12} + w_2 \sigma_2^2) \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \\
&= w_1^2 \sigma_1^2 + w_1 w_2 \sigma_{21} + w_1 w_2 \sigma_{12} + w_2^2 \sigma_2^2 \\
&= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_{12} \geq 0
\end{aligned}
$$

since $\sigma_{12} = \rho_{12} \sigma_1 \sigma_2 \geq -\sigma_1 \sigma_2$.

Definition 8.1 formalizes the notion of a portfolio lying on the efficient frontier. Note that every portfolio is ultimately defined by the weights that determine its composition.

Definition 8.1 A frontier portfolio is one that displays minimum variance among all feasible portfolios with the same $E(\tilde{r}_p)$.

A portfolio p, characterized by w_p, is a frontier portfolio, if and only if w_p solves.[4]

[4] The problem below in vector notation is problem (QP) of Chapter 6.

$$\min_{w} \frac{1}{2} w^T V w$$

$$(\lambda)\ \text{s.t. } w^T e = E \left(\sum_{i=1}^{N} w_i E(\tilde{r}_i) = E(\tilde{r}_p) = E \right)$$

$$(\gamma)\ w^T \mathbf{1} = 1 \quad \left(\sum_{i=1}^{N} w_i = 1 \right)$$

where the superscript T stands for *transposed* (i.e., transforms a column vector into a row vector) and reciprocally, e denotes the column vector of expected returns to the N assets, $\mathbf{1}$ represents the column vector of ones, and λ, γ are Lagrange multipliers. Short sales are permitted (no nonnegativity constraints are present). The solution to this problem can be characterized as the solution to $\min_{\{w,\lambda,\gamma\}} L$, where L is the Lagrangian:

$$L = \frac{1}{2} w^T V w + \lambda(E - w^T e) + \gamma(1 - w^T \mathbf{1}) \tag{8.8}$$

Under these assumptions, the optimal w_p, λ and γ must satisfy Eqs. (8.9) through (8.11), which are the necessary and sufficient first-order conditions (FOCs):

$$\frac{\partial L}{\partial w} = Vw - \lambda e - \gamma \mathbf{1} = 0 \tag{8.9}$$

$$\frac{\partial L}{\partial \lambda} = E - w^T e = 0 \tag{8.10}$$

$$\frac{\partial L}{\partial \gamma} = 1 - w^T \mathbf{1} = 0 \tag{8.11}$$

In the lines that follow, we manipulate these equations to provide an intuitive characterization of the optimal portfolio proportions (8.17). From Eq. (8.9), $Vw_p = \lambda e + \gamma \mathbf{1}$, or

$$w_p = \lambda V^{-1} e + \gamma V^{-1} \mathbf{1}, \text{ and} \tag{8.12}$$

$$e^T w_p = \lambda(e^T V^{-1} e) + \gamma(e^T V^{-1} \mathbf{1}) \tag{8.13}$$

Since $e^T w_p = w_p^T e$, we also have, from Eq. (8.10), that

$$E(\tilde{r}_p) = \lambda(e^{\mathrm{T}} V^{-1} e) + \gamma(e^{\mathrm{T}} V^{-1} \mathbf{1}) \tag{8.14}$$

From Eq. (8.12), we have

$$\begin{aligned} \mathbf{1}^T w_p = w_p^T \mathbf{1} &= \lambda(\mathbf{1}^T V^{-1} e) + \gamma(\mathbf{1}^T V^{-1} \mathbf{1}) \\ &= 1 \text{ (by Eq. (8.10))} \\ 1 &= \lambda(\mathbf{1}^T V^{-1} e) + \gamma(\mathbf{1}^T V^{-1} \mathbf{1}) \end{aligned} \tag{8.15}$$

Notice that Eqs. (8.14) and (8.15) are two *scalar* equations in the unknowns λ and γ (since such terms as $e^T V^{-1} e$ are pure numbers!). Solving this system of two equations in two unknowns, we obtain

$$\lambda = \frac{CE - A}{D} \quad \text{and} \quad \gamma = \frac{B - AE}{D} \tag{8.16}$$

where

$$\begin{aligned} A &= \mathbf{1}^T V^{-1} e = e^T V^{-1} \mathbf{1} \\ B &= e^T V^{-1} e > 0 \\ C &= \mathbf{1}^T V^{-1} \mathbf{1} \\ D &= BC - A^2 \end{aligned}$$

Here we have used the fact that the inverse of a positive definite matrix is itself positive definite. It can be shown that D is also strictly positive. Substituting Eqs. (8.16) into Eq. (8.12), we obtain

$$\begin{aligned} w_p &= \underbrace{\underbrace{\frac{CE - A}{D}}_{\lambda}}_{scalar} \underbrace{V^{-1} e}_{vector} + \underbrace{\underbrace{\frac{B - AE}{D}}_{\gamma}}_{scalar} \underbrace{V^{-1} \mathbf{1}}_{vector} \\ &= \frac{1}{D}[B(V^{-1}\mathbf{1}) - A(V^{-1}e)] + \frac{1}{D}[C(V^{-1}e) - A(V^{-1}\mathbf{1})]E \\ w_p &= \underbrace{g}_{vector} + \underbrace{h}_{vector} \underbrace{E}_{scalar} \end{aligned} \tag{8.17}$$

Since the FOCs (Eqs. (8.9) through (8.11)) are a necessary and sufficient characterization for w_p to represent a frontier portfolio with expected return equal to E, any frontier portfolio can be represented by Eq. (8.17). This is a very nice expression; pick the desired expected return E and it straightforwardly gives the weights of the corresponding frontier portfolio with E as its expected return. The portfolio's variance follows as $\sigma_p^2 = w_p^T V w_p$, which is also straightforward. Efficient portfolios are those for which E exceeds the expected return on the minimum risk, risky portfolio. Our characterization thus applies to efficient portfolios as well: Pick an efficient E and Eq. (8.17) gives its exact composition. See Appendix 8.2 for an example.

Can we further identify the vectors g and h in Eq. (8.17)? In particular, do they somehow correspond to the weights of easily recognizable portfolios? The answer is positive. Since, if $E = 0$, $g = w_p$, g then represents the weights that define the frontier portfolio with $E(\tilde{r}_p) = 0$. Similarly, $g + h$ corresponds to the weights of the frontier portfolio with $E(\tilde{r}_p) = 1$, since $w_p = g + hE(\tilde{r}_p) = g + h\mathbf{1} = g + h$.

The simplicity of the relationship in Eq. (8.17) allows us to make two claims.

Proposition 8.1 The entire set of frontier portfolios can be generated by (are affine combinations of) g and $g + h$.

Proof To see this, let q be an arbitrary frontier portfolio with $E(\tilde{r}_q)$ as its expected return. Consider portfolio weights (proportions) $\pi_g = 1 - \mathrm{E}(\tilde{r}_q)$ and $\pi_{g+h} = \mathrm{E}(\tilde{r}_q)$; then, as asserted,

$$[1 - E(\tilde{r}_q)]g + E(\tilde{r}_q)(g + h) = g + hE(\tilde{r}_q) = w_q$$

The prior remark is generalized in Proposition 8.2.

Proposition 8.2 The portfolio frontier can be described as affine combinations of any two frontier portfolios, not just the frontier portfolios g and $g + h$.

Proof To confirm this assertion, let p_1 and p_2 be any two distinct frontier portfolios; since the frontier portfolios are different, $E(\tilde{r}_{p_1}) \neq E(\tilde{r}_{p_2})$. Let q be an arbitrary frontier portfolio, with expected return equal to $E(\tilde{r}_q)$. Since $E(\tilde{r}_{p_1}) \neq E(\tilde{r}_{p_2})$, there must exist a unique number α such that

$$E(\tilde{r}_q) = \alpha E(\tilde{r}_{p_1}) + (1 - \alpha)E(\tilde{r}_{p_2}) \tag{8.18}$$

Now consider a portfolio of p_1 and p_2 with weights α, $1-\alpha$, respectively, as determined by Eq. (8.18). We must show that $w_q = \alpha w_{p_1} + (1 - \alpha)w_{p_2}$.

$$\begin{aligned}
\alpha w_{p_1} + (1 - \alpha)w_{p_2} &= \alpha[g + hE(\tilde{r}_{p_1})] + (1 - \alpha)[g + hE(\tilde{r}_{p_2})] \\
&= g + h[\alpha E(\tilde{r}_{p_1}) + (1 - \alpha)E(\tilde{r}_{p_2})] \\
&= g + hE(\tilde{r}_q) \\
&= w_q, \ \text{since } q \text{ is a frontier portfolio}
\end{aligned}$$

What does the set of frontier portfolios, which we have characterized so conveniently, look like? Can we identify, in particular, the minimum variance portfolio? Locating that portfolio is surely key to a description of the set of all frontier portfolios. Fortunately, given our results thus far, the task is straightforward.

For any portfolio on the frontier,

$\sigma^2(\tilde{r}_p) = [g + hE(\tilde{r}_p)]^T V[g + hE(\tilde{r}_p)]$, with g and h as defined earlier.

Multiplying all this out (very messy) yields:

$$\sigma^2(\tilde{r}_p) = \frac{C}{D}\left(E(\tilde{r}_p) - \frac{A}{C}\right)^2 + \frac{1}{C} \tag{8.19}$$

where A, C, and D are the constants defined earlier. We can immediately identify the following: since $C > 0$, $D > 0$,

i. the expected return of the minimum variance portfolio is A/C;
ii. the variance of the minimum variance portfolio is given by $1/C$;
iii. Equation (8.19) is the equation of a parabola with vertex $(1/C, A/C)$ in the expected return/variance space and of a hyperbola in the expected return/standard deviation space. See Figures 8.3 and 8.4.

The extended shape of this set of frontier portfolios is due to the allowance for short sales as underlined in Figure 8.5.

What has been accomplished thus far? First and foremost, we have a much richer knowledge of the set of frontier portfolios: given a level of desired expected return, we can easily identify the relative proportions of the constituent assets that must be combined to

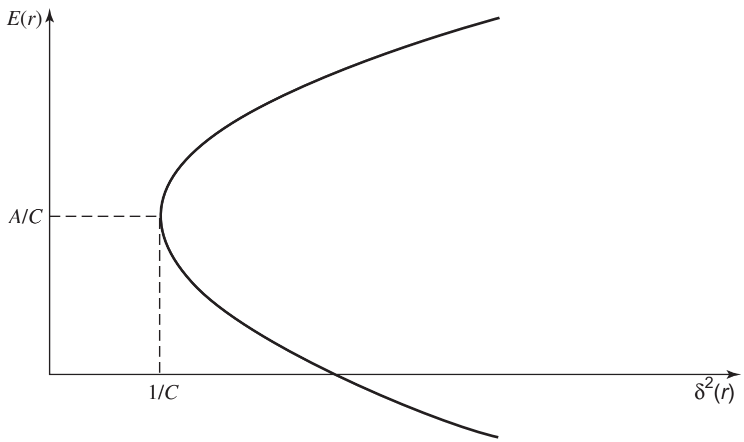

Figure 8.3
The set of frontier portfolios: Mean–variance space.

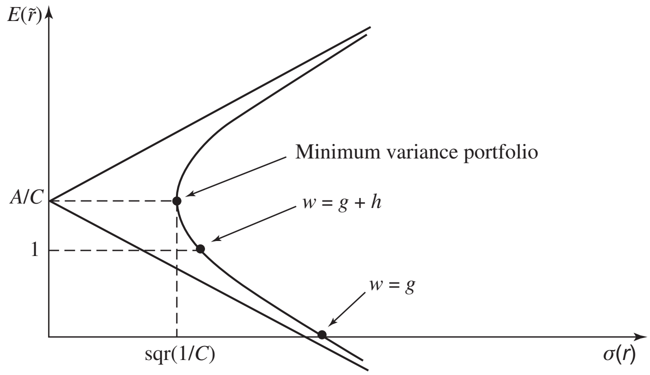

Figure 8.4
The set of frontier portfolios: Mean—standard deviation space.

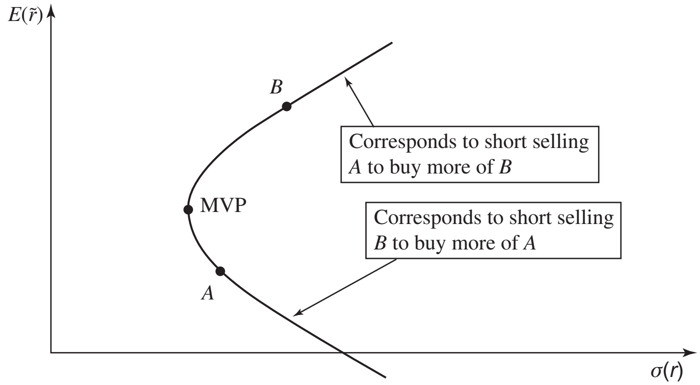

Figure 8.5
The set of frontier portfolios: Short selling allowed.

create a portfolio with that expected return. This was illustrated in Eq. (8.17), and it is key. We then used it to identify the minimum risk portfolio and to describe the graph of all frontier portfolios.

All of these results apply to portfolios of any arbitrary collection of assets. So far, nothing has been said about financial market equilibrium. As a next step toward that goal, however, we need to identify the set of frontier portfolios that is efficient. Given Eq. (8.17), this is a straightforward task.

8.5 Characterizing Efficient Portfolios (No Risk-Free Assets)

Our first order of business is a definition.

Definition 8.2 Efficient portfolios are those frontier portfolios for which the expected return exceeds A/C, the expected return of the minimum variance portfolio.

Since Eq. (8.17) applies to all frontier portfolios, it applies to efficient ones as well. Fortunately, we also know the expected return on the minimum variance portfolio. As a first step, let us prove the converse of Proposition 8.2.

Proposition 8.3 Any convex combination of frontier portfolios is also a frontier portfolio.

Proof Let $(\overline{w}_1 \ldots \overline{w}_N)$ define N frontier portfolios ($\overline{w}_i$ represents the vector defining the composition of the ith portfolio) and α_i, $i =, \ldots, N$ be real numbers such that $\sum_{i=1}^{N} \alpha_i = 1$. Lastly, let $E(\tilde{r}_i)$ denote the expected return of the portfolio with weights $\overline{w}_i$.

We want to show that $\sum_{i=1}^{N} \alpha_i \overline{w}_i$ is a frontier portfolio with $E(\overline{r}) = \sum_{i=1}^{N} \alpha_i E(\tilde{r}_i))$.

The weights corresponding to a linear combination of the above N portfolios are:

$$\begin{aligned}
\sum_{i=1}^{N} \alpha_i \overline{w}_i &= \sum_{i=1}^{N} \alpha_i (g + hE(\tilde{r}_i)) \\
&= \sum_{i=1}^{N} \alpha_i g + h \sum_{i=1}^{N} \alpha_i E(\tilde{r}_i) \\
&= g + h \left[\sum_{i=1}^{N} \alpha_i E(\tilde{r}_i) \right]
\end{aligned}$$

Thus $\sum_{i=1}^{N} \alpha_i \overline{w}_i$ is a frontier portfolio with $E(\tilde{r}) = \sum_{i=1}^{N} \alpha_i E(\tilde{r}_i)$.

A corollary to the previous result is the following preposition.

Proposition 8.4 The set of efficient portfolios is a convex set.[5]

Proof Suppose each of the N portfolios under consideration was efficient; then $E(\tilde{r}_i) \geq A/C$, for every portfolio i. However, $\sum_{i=1}^{N} \alpha_i E(\tilde{r}_i) \geq \sum_{i=1}^{N} \alpha_i A/C = A/C$; thus, the convex combination is efficient as well. So the set of efficient portfolios, *as characterized by their portfolio weights*, is a convex set.

[5] This does not mean, however, that the frontier of this set is convex-shaped in the risk-return space.

It follows from Proposition 8.4 that if every investor holds an efficient portfolio, the market portfolio, being a weighted average of all individual portfolios, is also efficient. This is a key result.

The next section further refines our understanding of the set of frontier portfolios and, more especially, the subset of them that is efficient. Observe, however, that as yet we have said nothing about equilibrium.

8.6 Background for Deriving the Zero-Beta CAPM: Notion of a Zero-Covariance Portfolio

Proposition 8.5 For any frontier portfolio p, except the minimum variance portfolio, there exists a unique frontier portfolio with which p has zero covariance.

We will call this portfolio the *zero-covariance portfolio relative to p* and denote its vector of portfolio weights by $ZC(p)$.

Proof To prove this claim it will be sufficient to exhibit the (unique) portfolio that has this property. As we shall demonstrate shortly (see Eq. (8.25) and the discussion following it), the covariance of any two frontier portfolios p and q is given by the following general formula:

$$\mathrm{cov}(\tilde{r}_p, \tilde{r}_q) = \frac{C}{D}\left[E(\tilde{r}_p) - \frac{A}{C}\right]\left[E(\tilde{r}_q) - \frac{A}{C}\right] + \frac{1}{C} \tag{8.20}$$

where A, C, and D are uniquely defined by e, the vector of expected returns, and V, the matrix of variances and covariances for portfolio p. These are, in fact, the same quantities A, C, and D defined earlier. If it exists, $ZC(p)$ must therefore satisfy

$$\mathrm{cov}(\tilde{r}_p, \tilde{r}_{ZC(p)}) = \frac{C}{D}\left[E(\tilde{r}_p) - \frac{A}{C}\right]\left[E(\tilde{r}_{ZC(p)}) - \frac{A}{C}\right] + \frac{1}{C} = 0 \tag{8.21}$$

Since A, C, and D are all numbers, we can solve for $E(\tilde{r}_{ZC(p)})$

$$E(\tilde{r}_{ZC(p)}) = \frac{A}{C} - \frac{\frac{D}{C^2}}{E(\tilde{r}_p) - \frac{A}{C}} \tag{8.22}$$

Given $E(\tilde{r}_{ZC(p)})$, we can use Eq. (8.17) to uniquely define the portfolio weights corresponding to it.

From Eq. (8.22), since $A > 0$, $C > 0$, $D > 0$, if $E(\tilde{r}_p) > A/C$ (i.e., is efficient), then $E(r_{Z\tilde{C}(p)}) < A/C$ (i.e., is inefficient), and vice versa. The portfolio $ZC(p)$ will turn out to be crucial to what follows. It is possible to give a more complete geometric identification to the zero-covariance portfolio if we express the frontier portfolios in the context of the $E(\tilde{r}) - \sigma^2(\tilde{r})$ space (Figure 8.6).

The equation of the line through the chosen portfolio p and the minimum variance portfolio can be shown to be the following (it has the form $(y = b + mx)$):

$$E(\tilde{r}) = \frac{A}{C} - \frac{\frac{D}{C^2}}{E(\tilde{r}_p) - \frac{A}{C}} + \frac{E(\tilde{r}_p) - \frac{A}{C}}{\sigma^2(\tilde{r}_p) - \frac{1}{C}}\sigma^2(\tilde{r})$$

If $\sigma^2(\tilde{r}) = 0$, then

$$E(\tilde{r}) = \frac{A}{C} - \frac{\frac{D}{C^2}}{E(\tilde{r}_p) - \frac{A}{C}} = E(\tilde{r}_{ZC(p)})$$

(by Eq. (8.22)).

That is, the intercept of the line joining p and the minimum variance portfolio is the expected return on the zero-covariance portfolio. This identifies the zero-covariance portfolio to p geometrically. We already know how to determine its precise composition.

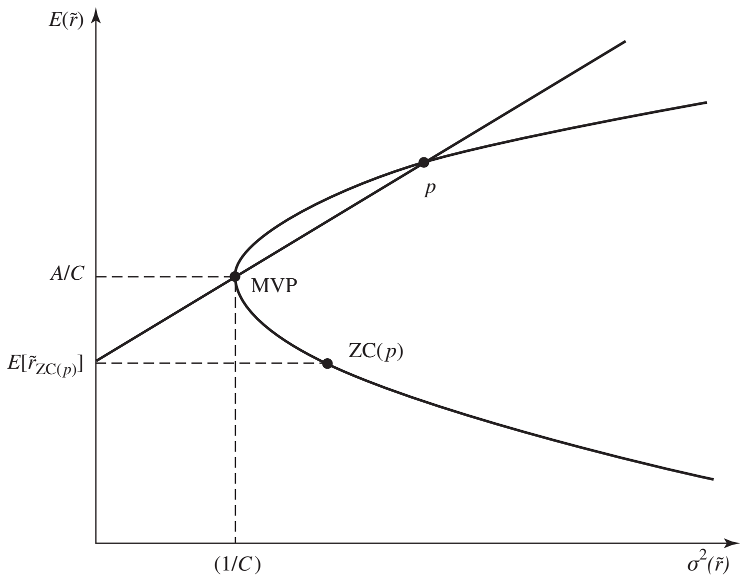

Figure 8.6
The set of frontier portfolios: Location of the zero-covariance portfolio.

Our next step is to describe the expected return on any portfolio in terms of frontier portfolios. After some manipulations, this will yield Eq. (8.29). The specialization of this relationship will give the zero-beta CAPM, which is a version of the CAPM when there is no risk-free asset. Recall that thus far we have not included a risk-free asset in our collection of assets from which we construct portfolios. Let q be any portfolio (which might not be on the portfolio frontier) and let p be any frontier portfolio.

$$\begin{aligned}\mathrm{cov}(\tilde{r}_p, \tilde{r}_q) &\underbrace{=}_{\text{by definition}} w_p^T V w_q \\ &= [\lambda V^{-1} e + \gamma V^{-1} \mathbf{1}]^T V w_q \\ &= \lambda e^T V^{-1} V w_q + \gamma \mathbf{1}^T V^{-1} V w_q \\ &= \lambda e^T w_q + \gamma \left(\text{since} \mathbf{1}^T w_q = \sum_{i=1}^N w_q^i \equiv 1 \right)\end{aligned} \tag{8.23}$$

$$= \lambda E(\tilde{r}_q) + \gamma \left(\text{since } e^T w_q = \sum_{i=1}^N E(\tilde{r}_i) w_q^i \equiv E(\tilde{r}_q) \right) \tag{8.24}$$

where $\lambda = (CE(\tilde{r}_p) - A)/D$ and $\gamma = (B - AE(\tilde{r}_p))/D$, as per earlier definitions.

Substituting these expressions into Eq. (8.24) gives

$$\mathrm{cov}(\tilde{r}_p, \tilde{r}_q) = \frac{CE(\tilde{r}_p) - A}{D} E(\tilde{r}_q) + \frac{B - AE(\tilde{r}_p)}{D} \tag{8.25}$$

Equation (8.25) is a short step from Eq. (8.20): Collect all terms involving expected returns, add and subtract A^2C/DC^2 to get the first term in Eq. (8.20) with a remaining term equal to $(1/C)(BC/D - A^2/D)$. But the latter is simply $1/C$ since $D = BC–A^2$.

Let us go back to Eq. (8.24) and apply it to the case where q is $ZC(p)$; one gets

$$0 = \mathrm{cov}(\tilde{r}_p, \tilde{r}_{ZC(p)}) = \lambda E(\tilde{r}_{ZC(p)}) + \gamma \text{ or } \gamma = -\lambda E(\tilde{r}_{ZC(p)}) \tag{8.26}$$

hence Eq. (8.24) becomes

$$\mathrm{cov}(\tilde{r}_p, \tilde{r}_q) = \lambda[\mathrm{E}(\tilde{r}_q) - \mathrm{E}(\tilde{r}_{ZC(p)})] \tag{8.27}$$

Apply the latter to the case $p = q$ to get

$$\sigma_p^2 = \mathrm{cov}(\tilde{r}_p, \tilde{r}_p) = \lambda[E(\tilde{r}_p) - E(\tilde{r}_{ZC(p)})] \tag{8.28}$$

and divide Eq. (8.27) by Eq. (8.28) and rearrange to obtain

$$E(\tilde{r}_q) = E(\tilde{r}_{ZC(p)}) + \beta_{pq}[E(\tilde{r}_p) - E(\tilde{r}_{ZC(p)})] \tag{8.29}$$

This equation bears more than a passing resemblance to the SML implication of the CAPM. But as yet it is simply a statement about the various portfolios that can be created from arbitrary collections of assets: (1) pick any frontier portfolio p; (2) this defines an associated zero-covariance portfolio $ZC(p)$; and (3) any other portfolio q's expected return can be expressed in terms of the returns to those portfolios and the covariance of q with the arbitrarily chosen frontier portfolio. Equation (8.29) would *very closely* resemble the SML if, in particular, we could choose $p = M$, the market portfolio of existing assets. The circumstances under which it is possible to do this form the subject to which we now turn.

8.7 The Zero-Beta CAPM

We would like to explain asset expected returns *in equilibrium*. The relationship in Eq. (8.29), however, is not the consequence of an equilibrium theory because it was derived for a *given* particular vector of expected asset returns, e, and a given covariance−variance matrix, V. In fact, it is the vector of returns e that we would like, in equilibrium, to understand. We need to identify a particular portfolio as being a frontier portfolio without specifying *a priori* the (expected) return vector and variance−covariance matrix of its constituent assets. The zero-beta CAPM tells us that under certain assumptions, this desired portfolio can be identified as the market portfolio M.

We may assume one of the following:

i. agents maximize expected utility with increasing and strictly concave utility of money functions, and asset returns are multivariate normally distributed, or
ii. each agent chooses a portfolio with the objective of maximizing a derived utility function of the form $W(e, \sigma^2)$, $W_1 > 0$, $W_2 < 0$, W concave.[6]

In addition, we assume that all investors have a common time horizon and homogeneous beliefs about e and V.

Under either set of assumptions, investors will only hold mean−variance efficient frontier portfolios.[7] But this implies that, **in equilibrium**, the market portfolio, which is a convex combination of individual portfolios, is also on the efficient frontier.[8]

6 As noted in Chapter 6, the maintained perspective of this text is alternative i. In this sense, the zero-beta version of the CAPM can be viewed as an implication of expected utility theory.

7 Recall the demonstration in Section 6.3.

8 Note that, in the standard version of the CAPM, the analogous claim crucially depended on the existence of a risk-free asset.

Therefore, in Eq. (8.23), p can be chosen to be M, the portfolio of all risky assets, and Eq. (8.29) can, therefore, be expressed as

$$E(\tilde{r}_q) = E(\tilde{r}_{ZC(M)}) + \beta_{\mathrm{M}q}[E(\tilde{r}_{\mathrm{M}}) - E(\tilde{r}_{ZC(\mathrm{M})})] \tag{8.30}$$

The relationship in Eq. (8.30) holds for any portfolio q, whether or not it is a frontier portfolio. This is the zero-beta CAPM.

An individual asset j is also a portfolio, so Eq. (8.30) applies to it as well:

$$E(\tilde{r}_j) = E(\tilde{r}_{ZC(M)}) + \beta_{\mathrm{M}j}[E(\tilde{r}_{\mathrm{M}}) - E(\tilde{r}_{ZC(\mathrm{M})})] \tag{8.31}$$

The zero-beta CAPM (and the more familiar *Sharpe−Lintner−Mossin* CAPM) is an equilibrium theory: the relationships in Eqs. (8.30) and (8.31) hold in equilibrium.[9] In equilibrium, investors will not be maximizing utility unless they hold efficient portfolios. Therefore, the market portfolio is efficient; we have identified one efficient frontier portfolio, and we can apply Eq. (8.30). By contrast, Eq. (8.29) is a pure mathematical relationship with no economic content; it simply describes relationships between frontier portfolio returns and the returns from any other portfolio of the same assets.

As noted in the introduction, the zero-beta CAPM does not, however, describe the process to or by which equilibrium is achieved. In other words, the process by which agents buy and sell securities in their desire to hold efficient portfolios, thereby altering security prices and thus expected returns, and requiring further changes in portfolio composition is not present in the model. When this process ceases and all agents are optimizing given the prevailing prices, then all will be holding efficient portfolios given the equilibrium expected returns e and covariance−variance matrix V. Thus M is also an efficient portfolio. The efficiency of M is a principal implication of the CAPM.

Since, in equilibrium, agents' desired holdings of securities coincide with their actual holdings, we can identify M as the actual portfolio of securities held in the marketplace. There are many convenient approximations to M—the S&P 500 index of stocks being the most popular in the United States. The usefulness of these approximations, which are needed to give empirical content to the CAPM, is, however, debatable, as discussed later in the chapter.

As a final remark, let us note that the name "zero-beta CAPM" comes from the fact that $\beta_{ZC(\mathrm{M}),\mathrm{M}} = \mathrm{cov}(\tilde{r}_M, \tilde{r}_{ZC(M)})/\sigma^2_{ZC(M)} = 0$, by construction of $ZC(M)$; in other words, the beta of $ZC(M)$ is zero.

[9] Sharpe (1964), Lintner (1965), and Mossin (1966).

8.8 The Standard CAPM

Our development thus far did not admit the option of a risk-free asset. We need to add this if we are to achieve the standard form CAPM. On a purely formal basis, of course, a risk-free asset has zero covariance with M and thus $r_{\mathrm{f}} = E(\tilde{r}_{ZC(M)})$. Hence we could replace $E(\tilde{r}_{ZC(M)})$ with r_{f} in Eq. (8.31) to obtain the standard representation of the CAPM, the SML. But this approach is not entirely appropriate since the derivation of Eq. (8.31) presumed the absence of any such risk-free asset.

More formally, the addition of a risk-free asset substantially alters the shape of the set of frontier portfolios in the $[E(\tilde{r}), \sigma(\tilde{r})]$ space. Let us briefly outline the development here, which closely resembles what was done above. Consider N risky assets with expected return vector e, and one risk-free asset, with *expected* return$\equiv r_{\mathrm{f}}$. Let p be a frontier portfolio, and let w_p denote the N vector of portfolio weights on the risky assets of p; w_p in this case is the solution to

$$\min_{w} \frac{1}{2} w^T V w$$

$$\text{s.t.} \;\; w^T e + (1 - w^T \mathbf{1}) r_{\mathrm{f}} = E$$

Solving this problem gives

$$w_p = V^{-1}(e - r_{\mathrm{f}}\mathbf{1}) \frac{E - r_{\mathrm{f}}}{H}$$

where $H = B - 2Ar_{\mathrm{f}} + Cr_{\mathrm{f}}^2$ and A, B, C are defined as before.

Let us examine this expression for w_p more carefully:

$$w_p = \underbrace{\underbrace{V^{-1}}_{nxn}\underbrace{(e - r_f\mathbf{1})}_{nx1}}_{nx1}\underbrace{\frac{E(\tilde{r}_p) - r_f}{H}}_{a\ number} \tag{8.32}$$

This expression tells us that if we wish to have a higher expected return, we should invest proportionally the same amount more in each risky asset so that the relative proportions of the risky assets remain unchanged. These proportions are defined by the $V^{-1}(e - r_{\mathrm{f}}\mathbf{1})$ term. This is exactly the result we were intuitively expecting: Graphically, we are back to the linear frontier represented in Figure 8.1.

The weights w_p uniquely identify the tangency portfolio T. Also,

$$\sigma^2(\tilde{r}_p) = w_p^T V w_p = \frac{[E(\tilde{r}_p) - r_{\mathrm{f}}]^2}{H}, \text{ and} \tag{8.33}$$

$$\mathrm{cov}(\tilde{r}_q, \tilde{r}_p) = w_q^T V w_p = \frac{[E(\tilde{r}_q) - r_\mathrm{f}][E(\tilde{r}_p) - r_\mathrm{f}]}{H} \tag{8.34}$$

for any portfolio q and any frontier portfolio p. Note how all this parallels what we did before. Solving Eq. (8.34) for $E(\tilde{r}_q)$ gives

$$E(\tilde{r}_q) - r_\mathrm{f} = \frac{H\ \mathrm{cov}(\tilde{r}_q, \tilde{r}_p)}{E(\tilde{r}_p) - r_\mathrm{f}} \tag{8.35}$$

Substituting for H via Eq. (8.33) yields

$$E(\tilde{r}_q) - r_\mathrm{f} = \frac{\mathrm{cov}(\tilde{r}_q, \tilde{r}_p)}{E(\tilde{r}_p) - r_\mathrm{f}} \frac{[E(\tilde{r}_p) - r_\mathrm{f}]^2}{\sigma^2(\tilde{r}_p)}$$

or

$$E(\tilde{r}_q) - r_\mathrm{f} = \frac{\mathrm{cov}(\tilde{r}_q, \tilde{r}_p)}{\sigma^2(\tilde{r}_p)} \left[E(\tilde{r}_p) - r_\mathrm{f}\right] \tag{8.36}$$

Again, since T is a frontier portfolio, we can choose $p \equiv T$. But in equilibrium $T = M$; in this case, Eq. (8.36) gives

$$E(\tilde{r}_q) - r_\mathrm{f} = \frac{\mathrm{cov}(\tilde{r}_q, \tilde{r}_M)}{\sigma^2(\tilde{r}_M)} [E(\tilde{r}_M) - r_\mathrm{f}]$$

or

$$E(\tilde{r}_q) = r_\mathrm{f} + \beta_{qM}[E(\tilde{r}_M) - r_\mathrm{f}] \tag{8.37}$$

for any asset (or portfolio) q. This is the standard CAPM.

Again, let us review the flow of logic that led to this conclusion. First, we identified the efficient frontier of risk-free and risky assets. This efficient frontier is fully characterized by the risk-free asset and a specific tangency frontier portfolio. The latter is identified in Eq. (8.32). We then observed that all investors, in equilibrium under homogeneous expectations, would hold combinations of the risk-free asset and that portfolio. Thus it must constitute the *market*—the portfolio of all risky assets. It is these latter observations that begin to give the CAPM empirical content.

8.9 An Empirical Assessment of the CAPM

Do observed patterns in financial data illustrate the conclusions of the CAPM? To answer this question, let us review its major assertions.[10] They are as follows:

1. Investors are well diversified; more precisely, investors hold the risky portion of their portfolio wealth in the form of units of the market portfolio M. Empirically, investors holding risky assets should be found to hold them in the form of some market index fund or otherwise very well-diversified portfolio. This means idiosyncratic risks are irrelevant.
2. A security's equilibrium expected return is a linear function of its beta, with the slope coefficient approximately equal to $[E\tilde{r}_M - r_f]$, and the intercept approximately equal to r_f. The empirical confirmation of this statement would be the observation that overextended periods of time, securities' average returns are linearly related to their respective betas with slope equal to the observed market risk premium, and intercept equal to the average risk-free rate (all measurements for the same data set). This is a statement about the cross section of security returns.
3. No other characteristics of a security matter for the determination of its average returns beyond the level of its systematic risk as measured by beta. Empirically this means that it should not be possible to find any other security characteristic (in addition to beta) that is useful in explaining the cross section of observed average security returns.

If we summarize the CAPM by the above three implications, the answer to the question posed at the start of this section is "no": the CAPM is an empirical failure.

In many respects, this failure is not at all surprising given the powerful assumptions underpinning the CAPM and their lack of empirical fulfillment. First and foremost, investors do not have identical expectations (the same forecasts of μ_i, σ_i, ρ_{ij}s for all risky assets under consideration). Second, the CAPM presumes that an investor's total wealth is derived only from his investments in the market portfolio and risk-free assets with the implication that *the market portfolio must contain every risky asset*. In reality, investors typically have wage income streams which are not perfectly positively correlated with $\tilde{r}_M$. These wage income streams are not tradable and thus are not part of the market portfolio. Furthermore, any convenient stock market proxy for the market portfolio, such as the $S\&P_{500}$ portfolio or the Wiltshire 5000 portfolio, inevitably omits important asset classes, most especially residential real estate and risky corporate debt. Forty percent of private non-human wealth in the United States takes the form of residential real estate. Taken together, these considerations imply that any empirical test of the CAPM will, in essence, be a joint test of the theory itself and the "efficiency" of the chosen market proxy, an observation

[10] What follows in this section is not a comprehensive review of the CAPM testing literature. Recent detailed surveys include Ferson and Jagannathan (1996) and Shanken (1996).

made forcefully by Roll (1977).[11] Lastly, many investors are constrained in their ability to borrow, which may lead them to form high-risk—high-expected return portfolios inside the efficient frontier.

With these qualifications in mind, let us now review the three CAPM summary conclusions in light of "the data."

The failure of the first assertion is immediate: many investors are not well diversified domestically or internationally, many hold large positions in their own firm's stock, and many own no stock at all (also a violation of the conclusion to Theorem 5.1), or only small amounts in the form of isolated individual shares.

To explore assertions (2) and (3) above, we first observe that they are basically statements about the form of the SML. Accordingly, it is this relationship that has been subject to formal statistical tests. We limit our review of these tests to three prominent papers and eschew any claim to being comprehensive. We also ignore the numerous details of data assembly and test design. These details are best addressed in a financial econometrics course. All three studies focus on the stock market where detailed data is readily available. The flow of the discussion is chronological.

8.9.1 Fama and MacBeth (1973)

These authors propose a two-step regression procedure, which has been used in numerous subsequent tests of the CAPM.[12] A simplified version of the essentials follows.

i. First-pass regression: Select a portfolio of assets to serve as the proxy for M. For each of the risky assets in the portfolio and the portfolio itself (value weighted), assemble its time series return data for some representative historical period; for discussion purposes here let this be quarterly data from 1966.1 through 2000.4.[13] For each asset i, estimate its historical beta via regression equation:

$$\tilde{r}_{i,s} - \tilde{r}_{f,s} = \hat{\alpha}_i + \hat{\beta}_i(\tilde{r}_{M,s} - \tilde{r}_{f,s}) + \tilde{\varepsilon}_{i,s} \tag{8.38}$$

   Equation (8.38) is simply the linear representation of the CAPM (cf. Eq. (8.5)). To emphasize the time series nature of this regression, we have assigned a subscript s

[11] Furthermore, modestly different proxies for M can sometimes lead to very different beta estimates for assets common to both.

[12] Fama and MacBeth (1973) were not the only authors to subject the CAPM to rigorous statistical tests. In particular, Friend and Blume (1970) for the 1960—1968 data period and Haugen and Heins (1975) using data from 1926 to 1971 provided robust evidence earlier on the limited ability of the CAPM to explain the data.

[13] In the case of Fama and MacBeth (1973), the proxy for the market portfolio is the set of all NYSE stocks traded sometime in the period 1926.1 through 1968.6, with returns computed monthly. The risk free rate $r_{f,t}$ is the monthly T-bill rate.

to all the variables. Since individual stock betas are often imprecisely estimated, and in order to minimize the effects of these estimation errors in the regressions to follow, Fama and MacBeth (1973), for example, partition ("sort") their universe of stocks into 20 portfolios based on their first-pass regression $\hat{\beta}_i s$, ranking them from lowest to highest.[14] Denote these portfolio betas as $\hat{\beta}^P_{j,t-1}$ where the added $t-1$ subscript is to be identified with the entire data period 1966.1 through 1970.4, and j indexes the portfolios.

ii. Second-pass regression: For each set of portfolios constructed sequentially in step 1, compute its average return $\overline{r}^P_{j,t}, j = 1, 2, \ldots, N = 20$ over a subsequent historical period, say the period 1971.1 through 1975.4. We index all quantities obtained from this second block of time series data with the subscript "t." The data points $(\overline{r}^P_{j,t}, \hat{\beta}^P_{j,t-1})$ then allow the estimation of the following second-pass, cross-sectional (the variation is indexed by j) regression:

$$\overline{r}^P_{j,t} = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t}\hat{\beta}^P_{j,t-1} + \hat{\gamma}_{2,t}(\hat{\beta}^P_{j,t-1})^2 + \hat{\gamma}_{3,t}s(\varepsilon^P_{j,t-1}) + \tilde{u}^P_{j,t} \tag{8.39}$$

The presence of the third term in the regression constitutes a simple test for nonlinearities in the return−systematic risk relationship across the collection of portfolios. The fourth term permits testing for evidence of an unsystematic risk-return relationship (thereby contradicting the first implication of the CAPM). In particular, $s(\varepsilon^P_{j,t-1})$ is the average of the $\sigma_{e_{i,t-1}}$ for all stocks i in the jth portfolio.

This two step process is then repeated over and over again using progressively updated blocks of data. To illustrate this updating, consider the next first-pass regression to be based on the period 1967.1 through 1972.4 with the second-pass regression based on 1972.1 through 1976.4. The overall result is to generate a time series of estimates $\{\hat{\gamma}_{0,t}, \hat{\gamma}_{1,t}, \hat{\gamma}_{2,t}, \hat{\gamma}_{3,t}\}$ corresponding to the succession of data periods.[15] Their procedure basically tests whether current betas have anything to say about future average returns.

The results of this large and carefully detailed study are as follows: First $\overline{\hat{\gamma}}_2$and $\overline{\hat{\gamma}}_3$ (time averages, e.g., $\hat{\gamma}_2 = 1/T\sum_{t=1}^{T}\hat{\gamma}_{2,t}$) are not statistically different from zero which is support for both linearity and the basic form of the SML. While $\overline{\hat{\gamma}}_1$ is positive, it is, however, too small relative to the average risk premium on the market. Essentially the average return/systematic risk relationship across the 20 portfolios is "too flat": the returns to low beta portfolios are too high relative to what the CAPM would predict while high beta portfolios' average returns are too low. Lastly, the $\overline{\hat{\gamma}}_0$ estimate is positive but

[14] In other words, the lowest beta portfolio is composed of the bottom 5% of stocks based on their beta rankings; the highest beta portfolio is composed of the highest 5% of stocks as measured by their estimated betas, etc. For each portfolio so composed, its portfolio beta is calculated as per $\beta_P = \sum_{k=1}^{N_P} w_k\hat{\beta}_k$, for all stocks k in the portfolio. In this way, the estimation errors are "averaged out."

[15] To make clear the timing of these regressions, see the diagrammatic representation in Appendix 8.3.

substantially in excess of the prevailing r_f. The Fama and MacBeth (1973) results have been replicated using other data sets covering different data periods. At best, their analysis represents a "lukewarm" confirmation of the CAPM.

8.9.2 Banz (1981) and the "Size Effect"[16]

The contribution of Banz (1981) is to initiate sorting on and to use explanatory variables based on characteristics unrelated to a stock's market beta. In particular, Banz (1981) double sorts all NYSE (New York Stock Exchange) traded stocks into 25 portfolios, first sorting on increasing "size"—defined to be the aggregate market value of a firm's equity (5 portfolio sort). In each size-ranked portfolio, the stocks are then "second-sorted" on the basis of increasing beta as per Fama and MacBeth (1973). In response to the criticism in Roll (1977), Banz (1981) also expands the market portfolio to include all publicly traded corporate and government debt. With these innovations, the Banz (1981) procedure amounts to the same as in Fama and MacBeth (1973). His "second-pass" regression takes the form

$$\bar{r}^P_{j,t} = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t}\hat{\beta}^P_{j,t-1} + \hat{\gamma}_{2,t}\left[\frac{ME^P_{j,t-1} - ME^P_{M,t-1}}{ME^P_{M,t-1}}\right] + \tilde{u}_{j,t} \tag{8.40}$$

In regression (8.40), $ME^P_{j,t}$ signifies the average market value of the stocks in portfolio j, while $ME^P_{M,t}$ is the average market value of all stocks in the market portfolio. Banz (1981) finds that the coefficients $\{\hat{\gamma}_{2,t}\}$ are on average negative and statistically significant, a finding in direct contradiction to the third assertion of the CAPM. While the "small firm effect" appears to have diminished in recent years, it nevertheless lives on in a somewhat transformed state (see the remarks on Fama and French (1993) in Chapter 14).

8.9.3 Fama and French (1992)

These authors cast a wider net for financial quantities with the power to explain the cross section of average equity returns. In particular, they evaluate BE/ME (see Section 2.5.6), E/P (the earnings to price ratio) and leverage (measured as A/ME, the book value of assets to market equity) in addition to the by-then-standard "size" (ME) and market-beta variables.

This selection has its basis in a number of earlier works. Stattman (1980) and Rosenberg et al. (1985) had found that firms' average equity returns were positively related to their BE/ME ratio. Chan et al. (1991) confirmed that BE/ME was also significant in explaining the cross section of average Japanese equity returns, while Basu (1983) had demonstrated that E/P provided added explanatory power for US average stock returns when included in tests involving both size and market-beta variables.

[16] See also Reinganum (1981) who simultaneously and independently discovered the "size effect."

In Section 2 of their paper, Fama and French (1992) undertake second-pass regressions using the explanatory variables under study both individually and in combinations thereof. In the case of *BE*/*ME*, they investigate the cross-sectional regression

$$\tilde{\bar{r}}_{i,t} = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t} \ \ln\left(\left(\widetilde{\frac{BE}{ME}}\right)_{i,t-1}\right) + \tilde{u}_{i,t} \tag{8.41}$$

using monthly data for nearly all NYSE stocks for the period 1963.1 through 1990.4.[17] Since $BE_{i,t}$ and $ME_{i,t}$ are measured precisely, there is no need to sort individual stocks into portfolios, and the entire panel of eligible stocks can be used with monthly updating. For regression (8.41), the time series average of the regression slopes is (0.50) with t-statistic (5.71). Clearly, the slope coefficient $\hat{\gamma}_1 = 1/T \sum_{t=1}^{T} \hat{\gamma}_{1,t}$ is both large and highly statistically different from zero.

This result is to be contrasted with the corresponding one for the market beta (with appropriate presorting into beta-ranked portfolios, etc.):

$$\bar{r}_{i,t} = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t}\hat{\beta}_{i,t-1} + \tilde{\varepsilon}_{i,t} \tag{8.42}$$

The average coefficient on beta is small (0.15) and its t-statistic (0.46) indicates that there is no statistical basis for concluding it is different from zero.

Fama and French (1992) offer three general conclusions to their study:

For the period 1963–1990:

1. "The market β does not seem to help to explain the cross section of average stock returns."
2. "The combination of size and book-to-market equity seems to absorb the roles of leverage and *E*/*P* in average stock returns."
3. The ratio of book-to-market equity is the most significant quantity for explaining the cross section of average security returns.[18]

While these observations leave us unsatisfied intellectually, it is clearly evident that the CAPM, however elegant and logical as a theory, fails to find empirical support.

8.9.4 Volatility Anomalies

CAPM theory, at its most basic, claims an *ex ante* positive relationship between an asset's undiversifiable risk and its expected returns: assets which have more undiversifiable risk

[17] Financial stocks are excluded from the sample. For each stock, $\bar{r}_{i,t}$ is the average monthly return for the 12 months following the end of period t.

[18] Both quotes from Fama and French (1992), p. 428.

(higher βs) are less desirable and should sell for lower prices and pay higher expected returns. In the cross section of security returns, the systematic risk (β) and total risk (σ) measures are statistically very highly correlated. The same risk–return relationship should thus hold whether risk is measured as total risk σ or as (relative) systematic risk β.[19] It is this claim that we presently consider.

Greenwood et al. (2010) (see also Ang et al. (2006, 2009)) explore the assertion using as their universe of stocks the Russell 1000, which they sort (with quarterly resorting) into quintiles based on total risk. Given this sorting, five equally weighted portfolios are formed, and their risks, returns, and betas measured using data for the period 1980–2009. The results are presented in Table 8.1 and graphed in Figure 8.7.

Table 8.1: Risk and return (annualized) quintiles of Russell 1000 stocks[a]

	Return	σ (Total Risk)	β_{CAPM}
Quintile 1 (low risk)	13.84%	12.15	0.62
2	14.58%	15.1%	0.88
3	15.16%	17.3%	1.02
4	12.89%	20.9%	1.25
5	8.06%	32.6%	1.74
Russell 1000 (M)	11.50%	15.6%	1.00

Source: Table 8.1 and the subsequent Figure 8.9 are taken from Greenwood et al. (2010).
[a]Returns and risks have been annualized from quarterly calculations; data period: (1980)–(2009); $E(\tilde{r})_{annual} = 12E(\tilde{r})_{monthly}$; $\sigma_{annual} = \sqrt{12}\sigma_{month}$. Recall our continuous compounding discussion in Chapter 3.

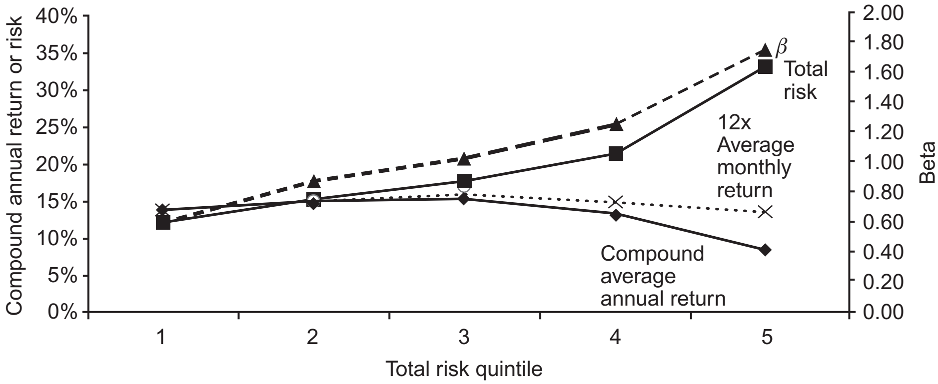

Figure 8.7
Graphical representation of Table 8.1.

[19] This result is not theoretically guaranteed. It is easy to imagine two stocks, 1 and 2, where $\beta_1 < \beta_2$ yet $\sigma_1 > \sigma_2$. See the definition of the CAPM β.

Table 8.1 merits a number of observations. First, we confirm that higher average β portfolios also display higher total risk: systematic risk and total risk go together at this quintile level of aggregation. In Table 8.1, note that $\sigma_M = 15.6\%$ is the risk of the market proxy, in this case the Russell 1000 stock portfolio.

Second, we note that up to and including the third quintile, the CAPM risk/return relationship is observed. For quintile portfolios four and five, however, this "intuitive" ordering is reversed: increasingly risky (by either risk measure) assets pay, on average, lower returns. One way of summarizing the pattern observed in Figure 8.7 is to say that the historical risk/return relationship is "too flat relative to the SML" (see also Fama and MacBeth, 1973). Not only is this phenomenon a violation of CAPM predictions, it appears also to conflict with the fundamental assumption of investor risk aversion.

There are several proposed explanations in the literature. Frazzini and Pederson (2013), for example, propose that the pattern in Figure 8.8 may reflect the fact that high-risk tolerant investors are subject to borrowing constraints (a violation of the CAPM assumptions): these investors would prefer to increase their expected returns in an efficient manner by leveraging the market portfolio proxy, but such leverage is unavailable to them. Frazzini and Pederson (2013) propose that this same investor class, as a second best alternative, will attempt to assemble portfolios composed exclusively of high beta, high expected return stocks. If a significant fraction of stock market investors act in this way, the equilibrium effect will be to increase the relative prices and decrease the average returns of high market-beta stocks relative to the CAPM prediction, as suggested by Figure 8.7. We emphasize that the absence of borrowing constraints is a fundamental assumption underlying the CAPM.

It has also been suggested that the risk and return pattern in Figure 8.7 may reflect high demand for what are called "lottery stocks." These are stocks of near-to-bankruptcy firms that have characteristics (low prices, tiny probabilities of gigantic upward price increases) similar to a state lottery ticket. Kumar (2009) classifies a stock as a lottery stock on the basis of low price, high idiosyncratic risk, and high idiosyncratic skewness. These are stocks that typically underperform given their betas, possibly contributing to the pattern of Figure 8.7. Kumar (2009) introduces evidence that the socioeconomic characteristics associated with those who purchase lottery tickets are also associated with the class of investors who purchase "lottery stocks," suggesting the possibility of the existence of a significant clientele that overweighs this type of security.

To characterize the purchasers of lottery tickets/lottery stocks simply as "risk lovers" would be, however, inaccurate. Rather, these persons identify themselves as individuals for whom the only way to accumulate significant wealth is to "hit the jackpot." They eschew, for whatever reasons or circumstances, the standard notion of wealth accumulation via savings and reinvestment. As such, their behavior is beyond the scope of expected utility theory cum risk aversion which underlies the CAPM analysis. We also note that lottery stocks have return distributions which are distinctly non-normal, and thus are formally excluded from consideration under the CAPM assumptions.

We recall from Chapter 2 that the idiosyncratic risk of most stocks represents 80−90% of their total return risk. Under the CAPM, idiosyncratic risks should not matter for asset pricing and average return determination: these are the risks that are diversified away. Emphasizing another departure from the assumptions of the CAPM, Merton (1987) and Hirshleifer (1988) show, via different mechanisms, that if investors face sizable trading frictions (and thus are not well diversified) and face incomplete markets, then idiosyncratic volatility should be positively linked with subsequent average returns.

In a pair of companion papers Ang et al. (2006, 2009) point out that for both US and international stocks, it is the reverse pattern that is actually observed. Measuring idiosyncratic volatility using the Fama−French three factor model (see Chapter 14), and sorting stocks into five idiosyncratic volatility portfolios, with quarterly rebalancing, Ang et al. (2006) uncover the reverse relationship most profoundly for high idiosyncratic volatility stocks. A similar relationship is observed if subsequent returns are compared across current period idiosyncratic risk rankings or if the returns and idiosyncratic risks are

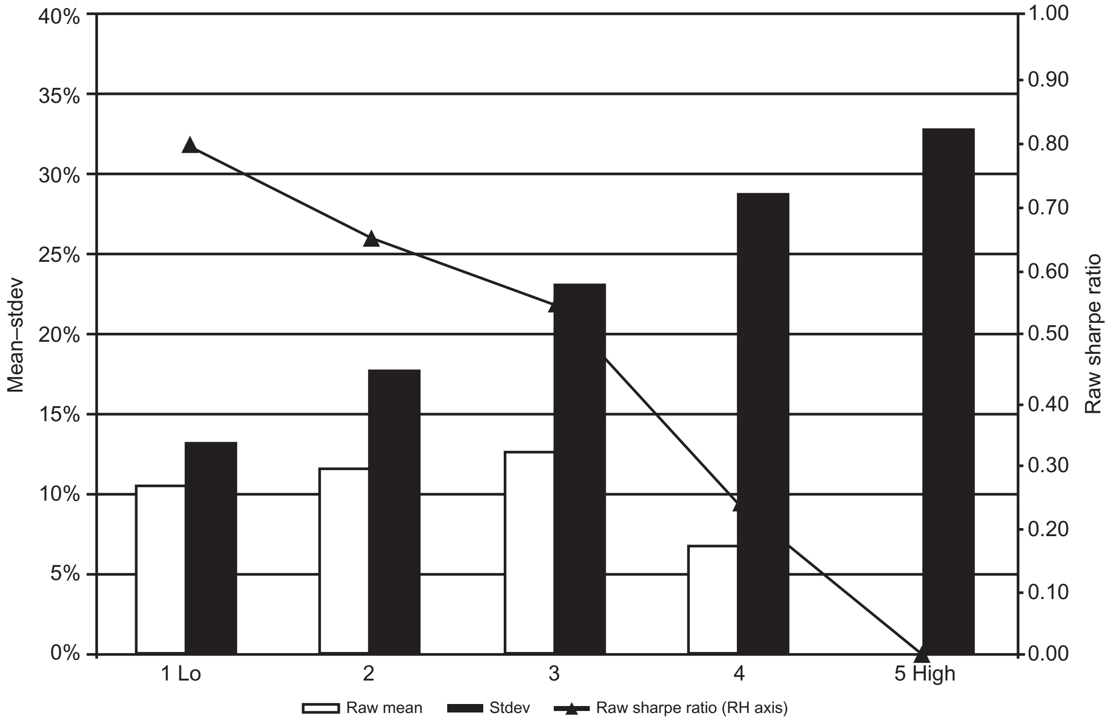

Figure 8.8
Average returns and idiosyncratic volatility: idiosyncratic volatility and subsequent returns.

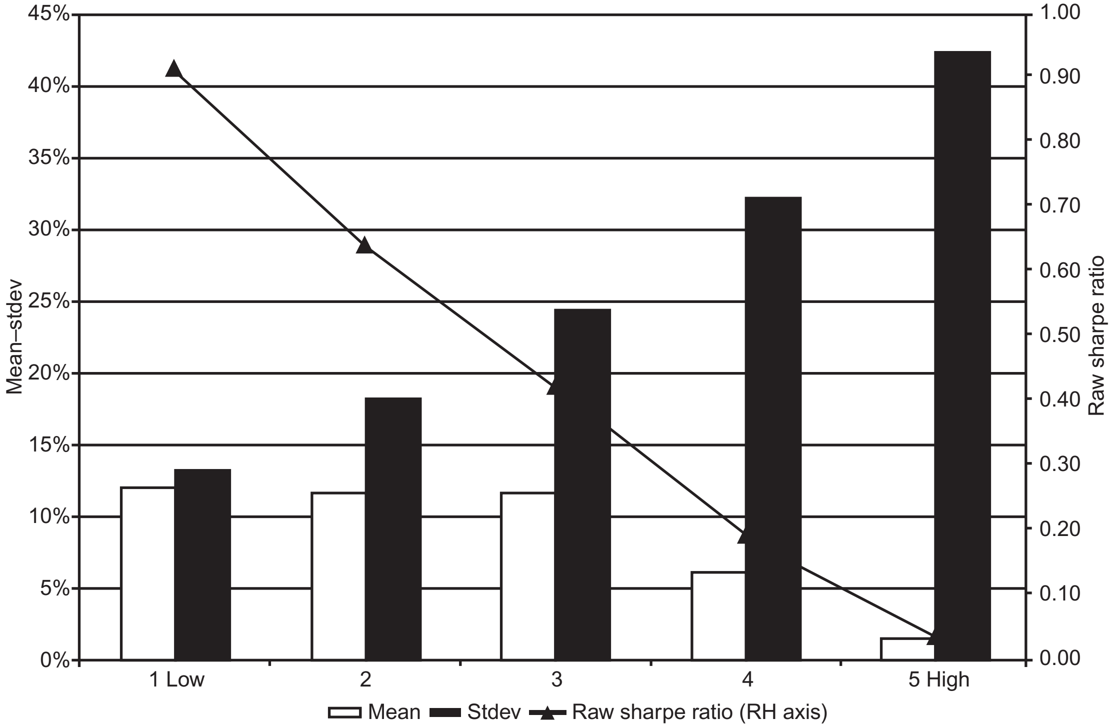

Figure 8.9
Average returns and idiosyncratic volatility contemporaneously measured.

measured contemporaneously. The results are presented, respectively, in Figures 8.8 and 8.9.

At the moment, there is no generally accepted explanation for these results, but they may also be related to the "borrowing constrained—lottery stock" nexus. The resulting mispricing should, however, be arbitraged away by short sellers, yet it is not. As emphasized most recently in Stambaugh et al. (2012), this failure must be laid at the feet of the various forms of short-selling constraints. Borrowing and short selling constraints thus appear to complement one another in allowing mispricing (at least as defined by the CAPM) to persist for extended periods.[20]

8.10 Conclusions

The CAPM was the first model to allow financial economists to organize their thoughts on the risk/return trade-off in a systematic way. While originally developed as a descriptive

[20] See Lamont (2004), D'Avolio (2002), and Hong and Sraer (2012) for various perspectives on "short selling constraints."

theory, it has normative implications as well: at a minimum individual investors should be well diversified. This means investing that portion of their wealth designated for risky assets in the best available approximation to the market portfolio. Only in this way can an investor hope to approximate the best risk/return trade-off available to him. This assertion represents solid intuition and there is really no questioning of its legitimacy as a general principle.

The principal conclusion of the CAPM, the SML, is not borne out in direct tests of the data, however. Given the severity of its assumptions and the difficulty in assembling a verifiably accurate approximation to the true M, this outcome is not totally unexpected. The CAPM is also challenged by various anomalies which take the form of other factors that marginalize the market excess return factor in explaining the pattern of excess security returns. Nevertheless, we are reminded that these anomalies have no meaning apart from the CAPM which provides the benchmark against which they are identified. The same must be said of the research accomplishments that have followed from them. As such the CAPM remains a *fundamental* device for organizing our thoughts about the relationship of risk and return.

References

Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. J. Finan. 61, 259−299.

Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2009. High idiosyncratic volatility and low returns: international and further U.S. evidence. J. Finan. Econ. 91, 1−23.

Banz, R., 1981. The relationship between return and market value of common stocks. J. Finan. Econ. 9, 3−18.

Basu, S., 1983. The relationship between earnings yield, market value, and the return for NYSE common stocks: further evidence. J. Finan. 43, 507−528.

Chan, L., Hamao, Y., Lakonishok, J., 1991. Fundamentals and stock returns in Japan. J. Finan. 46, 1739−1789.

D'Avolio, G., 2002. The market for borrowing stocks. J. Finan. Econ. 66, 271−306.

Fama, E., French, K., 1992. The cross section of expected stock returns. J. Finan. 47, 427−465.

Fama, E., MacBeth, J., 1973. Risk, return and equilibrium empirical tests. J. Polit. Econ. 81, 607−636.

Ferson, W.E., Jagannathan, R., 1996. Econometric evaluation of asset pricing models. In: Maddala, G.S., Rao, C.R. (Eds.), Statistical Methods in Finance, Handbook of Statistics. Amsterdam, North Holland, p. 14.

Frazzini, A., Pederson, L.H., 2013. Betting Against Beta, Working Paper, New York University.

Friend, I., Blume, M., 1970. The measure of portfolio performance under uncertainty. Am. Econ. Rev. 60, 561−575.

Greenwood, L., Viciera, L., Ang, A., Eysenbach, M., Jacques, B., 2010. Report on the Risk Anomaly, Martingale Asset Management.

Haugen, R.A., Heins, J., 1975. The risk and rate of return on financial assets: some old wine in new bottles. J. Finan. Quant. Anal. 10, 775−784.

Hirshleifer, D., 1988. Residual risk, trading costs and commodity futures risk premia. Rev. Finan. Stud. 1, 173−193.

Hong, H., Sraer, D. 2012. Speculative Betas, Working Paper, Princeton University.

Kumar, A., 2009. Who gambles in the stock market. J. Finan. 64, 1889−1933.

Lamont, O., 2004. Going Down Fighting: Short Sellers vs. Firms, Working Paper, Yale University.

Lintner, J., 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. 47 (1), 13–37.
Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica. 41, 867–887.
Merton, R., 1987. A simple model of capital market equilibrium with incomplete information. J. Finan. 2, 483–510.
Mossin, J., 1966. Equilibrium in a capital asset market. Econometrica. 34 (4), 768–783.
Reinganum, M., 1981. A new empirical perspective on the CAPM. J. Finan. Quant. Anal. 16, 439–462.
Roll, R., 1977. A critique of the asset pricing theory's test—Part I: on past and potential testability of the theory. J. Finan. Econ. 4, 129–176.
Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. J. Portf. Manag. 11, 9–17.
Shanken, J., 1996. Statistical methods in tests of portfolio efficiency: a synthesis". In: Maddala, G.S., Rao, C.R. (Eds.), Statistical Methods in Finance, Handbook of Statistics. Amsterdam, North Holland, p. 14.
Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. J. Finan. 19 (3), 425–442.
Stambaugh, R.F., Yu, J., Yuan, Y., 2012. Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle, NBER Working Paper #18560.
Stattman, D., 1980. Book values and stock returns, The Chicago MBA: A Journal of Selected Papers, 4, 25–45.

Appendix 8.1: Proof of the CAPM Relationship

Refer to Figure 8.1. Consider a portfolio with a fraction $1-\alpha$ of wealth invested in an arbitrary security j and a fraction α in the market portfolio.

$$E(\tilde{r}p) = aE(\tilde{r}_{\mathrm{M}}) + (1-\alpha)E(\tilde{r}_j)$$
$$\sigma_p^2 = \alpha^2\sigma_{\mathrm{M}}^2 + (1-\alpha)^2\alpha_j^2 + 2\alpha(1-\alpha)\sigma_{j\mathrm{M}}$$

As α varies we trace a locus that

- passes through M
  (and through j)
- cannot cross the CML (why?)
- hence must be tangent to the CML at M

$$\text{Tangency} = \frac{dE(\tilde{r}_p)}{d\sigma_p}\Big|_{\alpha=1} = \text{slope of the locus at } M = \text{slope of CML} = \frac{E(\tilde{r}_{\mathrm{M}}) - r_{\mathrm{f}}}{\sigma_M}$$

$$\frac{dE(\tilde{r}_p)}{d\sigma_p} = \frac{dE(\tilde{r}_p)/d\alpha}{d\sigma_p/d\alpha}$$
$$\frac{dE(\tilde{r}_p)}{d\alpha} = \bar{r}_M - \bar{r}_j$$
$$2\sigma_p\frac{d\sigma_p}{d\alpha} = 2\alpha\sigma_M^2 - 2(1-\alpha)\sigma_j^2 + 2(1-2\alpha)\sigma_{jM}$$
$$\frac{dE(\tilde{r}_p)}{d\sigma_p} = \frac{(E(\tilde{r}_{\mathrm{M}}) - E(\tilde{r}_j))\sigma_p}{\alpha\sigma_M^2 - (1-\alpha)\sigma_j^2 + (1-2\alpha)\sigma_{jM}}$$

$$\left.\frac{\mathrm{dE}(\tilde{r}_p)}{\mathrm{d}\sigma_p}\right|_{\alpha=1} = \frac{(\mathrm{E}(\tilde{r}_\mathrm{M}) - \mathrm{E}(\tilde{r}_j))\sigma_M}{\sigma_M^2 - \sigma_{jM}} = \frac{\mathrm{E}(\tilde{r}_\mathrm{M}) - \overline{r}_f}{\sigma_M}$$

$$(\mathrm{E}(\tilde{r}_\mathrm{M}) - \mathrm{E}(\tilde{r}_j)) = \frac{(\mathrm{E}(\tilde{r}_\mathrm{M}) - \overline{r}_\mathrm{f})(\sigma_M^2 - \sigma_{jM})}{\sigma_M^2}$$

$$(\mathrm{E}(\tilde{r}_\mathrm{M}) - \mathrm{E}(\tilde{r}_j)) = (\mathrm{E}(\tilde{r}_\mathrm{M}) - r_\mathrm{f})\left(1 - \frac{\sigma_{jM}}{\sigma_M^2}\right)$$

$$\mathrm{E}(\tilde{r}_j) = r_\mathrm{f} + (\mathrm{E}(\tilde{r}_\mathrm{M}) - r_\mathrm{f})\frac{\sigma_{jM}}{\sigma_M^2}$$

Appendix 8.2: The Mathematics of the Portfolio Frontier: An Example

$$\text{Assume } e = \begin{pmatrix} \mathrm{E}(\tilde{r}_1) \\ \mathrm{E}(\tilde{r}_2) \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}; V = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}, \text{ i.e., } \rho_{12} = \rho_{21} = -\tfrac{1}{2}$$

Therefore,

$$V^{-1} = \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

check:

$$\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}\begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{4}{3} - \frac{1}{3} & \frac{1}{3} - \frac{1}{3} \\ -\frac{4}{3} + \frac{4}{3} & -\frac{1}{3} + \frac{4}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A = 1^T V^{-1} e = (1 \quad 1)\begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \underbrace{\frac{4}{3} + \frac{1}{3}}_{\frac{5}{3}} & \underbrace{\frac{1}{3} + \frac{1}{3}}_{\frac{2}{3}} \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$= \frac{5}{3} + 2\left(\frac{2}{3}\right) = 3$$

$$B = e^T V^{-1} e = \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{4}{3} + \frac{2}{3} & \frac{1}{3} + \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 4$$

$$C = 1^T V^{-1} 1 = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{5}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{7}{3}$$

$$D = BC - A^2 = 4\left(\frac{7}{3}\right) - 9 = \frac{28}{3} - \frac{27}{3} = \frac{1}{3}$$

Now we can compute g and h:

1. $$g = \frac{1}{D}\left[B(V^{-1}1) - A(V^{-1}e)\right]$$

$$= \frac{1}{\frac{1}{3}}\left[4\begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} - 3\begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right]$$

$$= 3\left[4\begin{pmatrix} \frac{5}{3} \\ \frac{2}{3} \end{pmatrix} - 3\begin{pmatrix} \frac{6}{3} \\ \frac{3}{3} \end{pmatrix}\right] = 3\left[\begin{pmatrix} \frac{20}{3} \\ \frac{8}{3} \end{pmatrix} - \begin{pmatrix} \frac{18}{3} \\ \frac{9}{3} \end{pmatrix}\right]$$

$$= \left[\begin{pmatrix} 20 \\ 8 \end{pmatrix} - \begin{pmatrix} 18 \\ 9 \end{pmatrix}\right] = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

2. $$h = \frac{1}{D}\left[C(V^{-1}e) - A(V^{-1}1)\right]$$

$$= \frac{1}{\frac{1}{3}}\left[\frac{7}{3}\begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3\begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right]$$

$$= 3\left[\frac{7}{3}\begin{pmatrix}2\\1\end{pmatrix} - 3\begin{pmatrix}\frac{5}{3}\\\frac{2}{3}\end{pmatrix}\right] = 7\begin{pmatrix}2\\1\end{pmatrix} - 3\begin{pmatrix}\frac{5}{3}\\\frac{5}{3}\end{pmatrix} = \begin{pmatrix}14\\7\end{pmatrix} - \begin{pmatrix}15\\6\end{pmatrix} = \begin{pmatrix}-1\\1\end{pmatrix}$$

Check by recovering the two initial assets; suppose $E(\tilde{r}_p) = 1$:

$$\begin{pmatrix}w_1\\w_2\end{pmatrix} = \begin{pmatrix}2\\-1\end{pmatrix} + \begin{pmatrix}-1\\1\end{pmatrix}E(\tilde{r}_p) = \begin{pmatrix}2\\-1\end{pmatrix} + \begin{pmatrix}-1\\1\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix} \Rightarrow \text{OK}$$

suppose $E(\tilde{r}_p) = 2$:

$$\begin{pmatrix}w_1\\w_2\end{pmatrix} = \begin{pmatrix}2\\-1\end{pmatrix} + \begin{pmatrix}-1\\1\end{pmatrix}2 = \begin{pmatrix}2\\-1\end{pmatrix} + \begin{pmatrix}-2\\2\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix} \Rightarrow \text{OK}$$

The equation corresponding to Eq. (7.16) thus reads:

$$\begin{pmatrix}w_1^p\\w_2^p\end{pmatrix} = \begin{pmatrix}2\\-1\end{pmatrix} + \begin{pmatrix}-1\\1\end{pmatrix}E(\tilde{r}_p)$$

Let us compute the minimum variance portfolio for these assets.

$$E(\tilde{r}_{p,\text{min var}}) = \frac{A}{C} = \frac{9}{7}$$

$$\sigma^2(\tilde{r}_{p,\text{min var}}) = \frac{1}{C} = \frac{3}{7} < \min\{1, 4\}$$

$$w^p = \begin{pmatrix}2\\-1\end{pmatrix} + \begin{pmatrix}-1\\1\end{pmatrix}\frac{9}{7} = \begin{pmatrix}2\\-1\end{pmatrix} + \begin{pmatrix}-\frac{9}{7}\\\frac{9}{7}\end{pmatrix} = \begin{pmatrix}\frac{14}{7}\\-\frac{7}{7}\end{pmatrix} + \begin{pmatrix}-\frac{9}{7}\\\frac{9}{7}\end{pmatrix} = \begin{pmatrix}\frac{5}{7}\\\frac{2}{7}\end{pmatrix}$$

Let's check $\sigma^2(\tilde{r}_p)$ by computing it another way:

$$\sigma_p^2 = \begin{pmatrix}\frac{5}{7} & \frac{2}{7}\end{pmatrix}\begin{pmatrix}1 & -1\\-1 & 4\end{pmatrix}\begin{pmatrix}\frac{5}{7}\\\frac{2}{7}\end{pmatrix} = \begin{pmatrix}\frac{3}{7} & \frac{3}{7}\end{pmatrix}\begin{pmatrix}\frac{5}{7}\\\frac{2}{7}\end{pmatrix} = \tfrac{3}{7} \Rightarrow \text{OK}$$

Appendix 8.3: Diagrammatic Representation of the Fama–MacBeth Two-Step Procedure

Data period "$t-1$"	Data period "t" (subsequent (to $t-1$) block of data)
– Perform first-pass regressions to obtain individual stock $\hat{\beta}_{i,t-1}, \sigma_{\varepsilon_{i,t}}$. – Form portfolio assignments (sorting) – Compute $\hat{\beta}^{P}_{j,t-1}, s(\tilde{\varepsilon}^{P}_{i,t-1})$ for the sorted portfolios	– Use $\hat{\beta}^{P}_{j,t-1}, s(\tilde{\varepsilon}^{P}_{i,t-1})$ together with $\bar{r}^{P}_{j,t}$ to perform the second-pass regressions

– Repeat the entire process using updated data sets to obtain $\hat{\gamma}_{0,t+1}, \hat{\gamma}_{1,t+1}, \hat{\gamma}_{2,t+1}$, and $\hat{\gamma}_{3,t+1}$.

– Analyze time series $\{\hat{\gamma}_{0,t}, \hat{\gamma}_{1,t}, \hat{\gamma}_{2,t}, \hat{\gamma}_{3,t}\}$ for statistical significance