Chapter III. Basics of the Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) is the most popular model of the determination of expected returns on securities and other financial assets. It is considered to be an "asset pricing" model since, for a given exogenous expected payoff, the asset price can be backed out once the expected return is determined. Additionally, the expected return derived within the CAPM or any other asset pricing model may be used to discount future cash flows. These discounted cash flows then are added to determine an asset's price. So, even though the focus is on expected return, we will continue to refer to the CAPM as an asset pricing model.

1. Derivation and Interpretation of the CAPM Pricing Formula

The basic CAPM model assumes the existence of a risk free asset and we assume this in the current section. Thus, the frontier results of sections 3(c) and 3(d) of Chapter II apply.

(a) Algebra of the Portfolio Frontier

Consider the *perceived* means and covariances of the various risky assets and the choices of a particular individual investor. In the mean-variance framework, the individual minimizes with respect to portfolio shares $\boldsymbol{s_p}$ the variance (half of it really, to simplify the resulting first-order condition) of the portfolio return subject to the constraint of a given expected return μ_p:

(1) Minimize $\frac{1}{2}\, \boldsymbol{s_p}^{\boldsymbol{T}} \Sigma\, \boldsymbol{s_p}$,

(2) Subject to: $\boldsymbol{s_p}^{\boldsymbol{T}} (\boldsymbol{\mu} - r_f) = \mu_p - r_f$.

Sections 3(c) and 3(d) of Chapter II provide more detail on this efficient portfolio choice decision problem.[1] Using the Lagrangian method with muliplier λ constraint (2) and differentiating with respect to $\boldsymbol{s_p}^{\boldsymbol{T}}$ produces the following first-order condition:

(3) $\Sigma\, \boldsymbol{s_p}^* = \lambda\, (\boldsymbol{\mu} - r_f)$.

Equation (3) gives the efficient vector of portfolio shares $\boldsymbol{s_p}^*$ given the investor's perceived means and covariances of the available assets and a particular mean portfolio return μ_p.

The covariance $Cov(r_i, r_p) \equiv \sigma_{ip}$ between the (excess) returns of one individual asset (or portfolio) *i* and a frontier portfolio *p* (that is, given the assumption of a risk free asset, a portfolio that puts this investor on his perceived

[1] The concept of (mean-variance) efficient portfolio choice will be used frequently in this chapter. It is important to distinguish the concept of an *efficient portfolio* from the fundamentally different concept of an *efficient market* which we will discuss in Chapter V.

CML for mean return μ_p) is given as:

$$(4) \qquad \sigma_{ip} = \boldsymbol{s_i}^{\boldsymbol{T}} \Sigma \, \boldsymbol{s_p}^*,$$

where $\boldsymbol{s_i}^{\boldsymbol{T}}$ indicates a row vector with s_i at position *i* and zeros elsewhere (or any transposed vector of risky asset shares when *i* is a portfolio). Employing equation (3) yields:

$$(5) \qquad \sigma_{ip} = \lambda \, \boldsymbol{s_i}^{\boldsymbol{T}} (\boldsymbol{\mu} - r_f).$$

Using equation (5) for $i = p$ gives:

$$(6) \qquad \sigma_p^2 = \lambda (\mu_p - r_f).$$

Eliminating λ from equations (5) and (6) yields:

$$(7) \qquad \mu_i - r_f = \beta_{ip} (\mu_p - r_f) \,,$$

where $\beta_{ip} \equiv \sigma_{ip}/\sigma_p^2$.

From basic econometrics we know that we can always write equation (7) as:

$$(8) \qquad r_i = r_f + \beta_{ip} (r_p - r_f) + \varepsilon_{ip},$$

where $Cov(r_p, \varepsilon_{ip}) = 0$, and $E(\varepsilon_{ip}) = 0$. To prove equation (8), consider that, theoretically, any linear equation, such as equation (7), with one independent variable can be written as equation (8), with a slope $\beta_{ip} \equiv \sigma_{ip}/\sigma_p^2$ [since $\varepsilon_{ip} = r_i - r_f - \beta_{ip} (r_p - r_f)$ and $\beta_{ip} \equiv \sigma_{ip}/\sigma_p^2$ it is then easy to check that $Cov(r_p, \varepsilon_{ip}) = 0$], and intercept of r_f [taking expectations in equation (8) and using equation (7) then implies $E(\varepsilon_{ip}) = 0$]. In fact, this formulation may be found exactly by running a simple OLS regression between r_i and r_p - r_f .

It is important to note that the derivation of equations (7) and (8) is valid for the perceived opportunities of any individual investor in isolation. It is tautologically true for any investor and any asset based on the mean-variance assumption; it follows from the mathematics of the portfolio frontier.

(b) The Capital Asset Pricing Model and Its Assumptions

The investor-specific result of equations (7) and (8) required the following assumptions, categorized by the part of the decision problem that requires the assumption:

Objectives

1. Investor preferences display risk aversion and non-satiation, and are quadratic; or, if preferences are not quadratic, asset returns are multi-variate elliptically distributed.

Note that the condition of ellipticality is of course technically an assumption on the assets rather than on the objectives.

2. One-period model.

The investor is myopic, considering only the current period. The effect of changes in investment opportunities over time is ignored. This assumption will be relaxed when we consider dynamic asset pricing models in Chapters VIII and IX.

3. Only total consumption matters.

The investor's utility function includes overall consumption as its only argument. There is no direct utility of diversifying or holding particular securities. The composition of overall consumption is irrelevant. We will discuss in a later chapter the implications of allowing the investor to have non-homothetic preferences over different consumption goods (like housing and other consumption).

Note that assumption 2 together with assumption 3 implies that only end-of-period wealth matters to the investor. Assumption 1 implies that the investor has mean-variance preferences over wealth such that he likes higher mean wealth and dislikes higher standard deviation of wealth. For any initial level of wealth the mean-variance preferences over wealth imply, of course, directly mean-variance preferences over portfolio returns. Assumptions 1-3 are sufficient to posit equation (1) as the key objective: Minimize $\frac{1}{2}\, \boldsymbol{s_p}^{\boldsymbol{T}} \Sigma\, \boldsymbol{s_p}$, which is optimal for a given mean portfolio return.

Market Conditions

As a first step in describing the investment opportunities available to investors, markets for all assets are assumed to be perfect.

4. Perfect competition.

The investor takes the asset's price (and so the perceived mean return and standard deviation) as given.

5. Absence of frictions

No taxes (such as capital gains, dividend income, or (financial) sales); no transaction costs (such as a fixed transaction cost independent of purchase value); no regulations (such as those restricting trades); no short sales restrictions (unlimited short sales are allowed, and borrowing and lending rates are equal).

6. All assets owned by the investor are marketable

Slavery is possible: future labor–human capital–can be sold or bought; a residence may be sold without giving up residence. In parts of Chapter IV this assumption is dropped.

7. Information on any asset, if available, can be obtained without cost

Having an investor decide whether to purchase information on any individual asset would substantially complicate matters. Relaxing assumption 7 is considered in Chapter V.

Note that absence of market imperfections, the assumption of *perfect* markets, is different than the assumption of *complete* markets which we will run into later.

Investment Opportunities

8. The types of assets are given exogenously.

There is no consideration of, say, firms stepping into the market to provide assets that would be particularly attractive to investors. The supply side is suppressed.

9. Assets are perfectly divisible.

This is a simplifying assumption that is quite reasonable for financial assets, especially for assets traded on major exchanges.

10. A riskless asset exists.

One could argue that due to inflation risk (if no bond exists indexed to your consumption basket), an unknown investment horizon (are short-term or long-term bonds risky for you? This would depend on your liquidity needs which may change over time), changing investment opportunities (interest rates may be up or down at the end of the period), and catastrophic risk (a major war or natural disaster may make any government default) no truly risk free asset exists. Later in this chapter we consider the model if no such risk free asset exists.

Given assumption 1, all asset returns must be assumed to be elliptically distributed if we do not assume quadratic preferences. The assumptions on market conditions and investment opportunities together are sufficient for equation (2) of the model, $\boldsymbol{s_p^T}(\boldsymbol{\mu} - r_f) = \mu_p - r_f$, to hold. With the assumptions on preferences added that imply equation (1), the model derivation of equations (7) and (8) follows logically. Note that all means, variances, and covariances must be interpreted thus far as perceived by one individual investor. To apply the model uniformly and make it useful for positive economic analysis, we need to add two more classes of assumptions that limit the differences among investors and define equilibrium.

Investor homogeneity

The above assumptions imply a CML for an individual investor. They are also sufficient to yield normative investment advice and imply equations (7) and (8) based on the individual investor's expectations. In order to derive the Mutual Fund Theorem or to prove that the price of risk reduction and the investment opportunities are equivalent for all investors we need to make the following additional assumptions:

11. Homogeneous availability and interpretation of information.

No difference exists between informed and uninformed investors. The investment opportunities are viewed in the same way by all investors. This assumption is dropped in Chapter V.

12. Homogeneous access to investment opportunities.

Rules out situations where investors are credit constrained due to investor-specific characteristics; rules out differences among investors in different countries caused by, for instance, exchange rate fluctuations. The latter issue will be addressed in Chapter IV.

With assumptions 11 and 12 added we can now view equations (7) and (8) as holding for all investors.

Market Equilibrium

So far the assumptions have no bearing on equilibrium asset pricing. Equations (7) and (8) are solely the implications of the rational (efficient) portfolio choices of individual investors. The return on any asset *i*, as perceived by an individual investor, can be related to the risk free rate and the perceived return on any perceived frontier portfolio. We now add the final assumption and then continue to derive the basic CAPM formula

13. Market clearing.

Prices for all assets are assumed to move such that an exogenous quantity of each asset equals the aggregate demand for the asset.

First define *market wealth* as the aggregate level of wealth:

$$(9) \qquad \bar{w}_m \equiv \sum_{k=1}^{K} \bar{w}_k \,,$$

where the individual initial wealth of the K investors in the economy is summed to get (initial) market wealth. Consider next the aggregate quantity of any asset *i* held in equilibrium. This is given as the equilibrium market share of asset *i*, s_{im}, times market wealth. Since *in equilibrium all assets are held* it must be true for any asset that:

$$(10) \qquad \sum_{k=1}^{K} s_{ik}\bar{w}_k = s_{im}\bar{w}_m \quad \leftrightarrow \quad \sum_{k=1}^{K} s_{ik}(\bar{w}_k/\bar{w}_m) = s_{im} .$$

Thus, the portfolio consisting of the market shares of all risky assets (the *market portfolio*) is a convex combination of the portfolios of all individuals [convex since $\Sigma_k(\bar{w}_k/\bar{w}_m) = 1$ from equation (9)]. We know that, in equilibrium, and using assumptions 11 and 12, all individuals hold frontier portfolios (and, given the assumption of a risk free asset, are on the CML). Since a convex combination of a frontier (CML) portfolio is still a frontier (CML) portfolio we now know that the market portfolio is a frontier (CML) portfolio. In fact, since the market portfolio is defined as excluding the risk free asset, we know that it must be the tangency portfolio, which is the only portfolio on the CML with zero weight on the risk free asset.[2] We can summarize this argument in the following syllogism: all individuals hold their risky assets in the same frontier portfolio p; the aggregation of all individual risky portfolios yields the market portfolio m. Thus, p equals m. Accordingly, we replace equations (7) and (8) by:

$$(11) \qquad \mu_i - r_f = \beta_i(\mu_m - r_f) ,$$

where the subscript *m* indicates the market portfolio; $\beta_i \equiv Cov(r_i, r_m)/\sigma_m^2$.

$$(12) \qquad r_i = r_f + \beta_i(r_m - r_f) + \varepsilon_i ,$$

where $Cov(r_m, \varepsilon_i) = 0$, and $E(\varepsilon_i) = 0$. Equations (11) and (12) provide the standard CAPM formulas, in expected returns form and in market realization form.

(c) Interpretation of the CAPM formula

Figure (1) shows the *Securities Market Line*, displaying the expected return of asset *i*, μ_i, as a linear function of its market *beta*, β_i. The expected excess return of any asset $\mu_i - r_f$ can be viewed as the risk premium of the asset. It consists of two components: the expected *market risk premium*, $\mu_m - r_f$, and the asset-specific beta.

Beta measures the "volatility" of an asset's return as a standardized quantity of *covariance risk*, the ratio of the asset return's covariance with the market return divided by the variance of the market return. Why do we consider covariance risk rather than the total variance of the asset as a measure of risk? In a 1998 interview, Sharpe stated the following about risk: "[T]here's no reason to expect reward just for bearing risk. Otherwise, you'd make a lot of money in Las Vegas. If there's reward for risk, it's got to be special." Define ρ_i as the correlation coefficient between the return of asset i and the market return. Then, using the definitions of beta and correlation coefficient,

[2] The market portfolio may be viewed as, in principle, including the risk free asset. In general equilibrium, however, when borrowing liabilities are offset with lending assets, the net supply of the risk free asset is typically equal to zero.

(11) $\beta_i \equiv Cov(r_i, r_m)/\sigma_m^2 \equiv \rho_i\, \sigma_i / \sigma_m .$

The risk specific to asset *i* can now be interpreted as: that part of asset return risk that is correlated with the market (and normalized by dividing by the standard deviation of the market return).

We can write tautologically:

(12) $\sigma_i \equiv \rho_i\, \sigma_i + (1 - \rho_i)\sigma_i \;.$

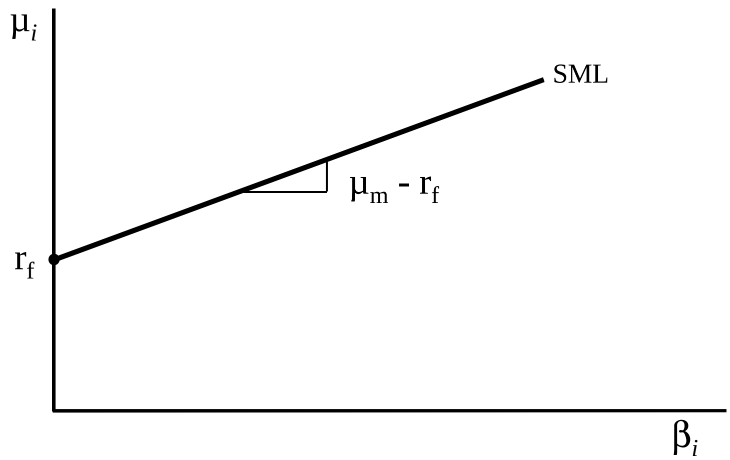

Figure 1
Security Market Line
The Security Market Line (SML) is the relationship between the expected return on asset *i* as a linear function of its market beta, β_i.

The first term on the right-hand side of equation (12) represents *systematic risk* the second term represents *idiosyncratic risk*. Only the systematic risk is valued in the CAPM context, the idiosyncratic risk is irrelevant for determining the expected return of an asset. The reason is that idiosyncratic risk can be averaged away in any well-diversified portfolio. The systematic risk, however, is unavoidable and should be priced. Thus beta provides a standardized measure of the relevant risk, systematic risk.

Note that other, more or less equivalent, names for idiosyncratic risk are: non-market risk, diversifiable risk, firm-specific risk, and non-systematic risk. The use of systematic or undiversifiable risk in this context is not quite correct and is a little confusing. It presumes a large portfolio such that all idiosyncratic risk is diversified away due to the *law of large numbers*. However, the covariance risk that is relevant in the CAPM is also defined when the market consists of only a few assets; the law of large numbers does not come into play as a motivation of the CAPM.

Alternatively, to interpret risk in the CAPM context, take the variance in equation (9):

(13) $\sigma_i^2 = \beta_i^2\, \sigma_m^2 + \sigma_\varepsilon^2 .$

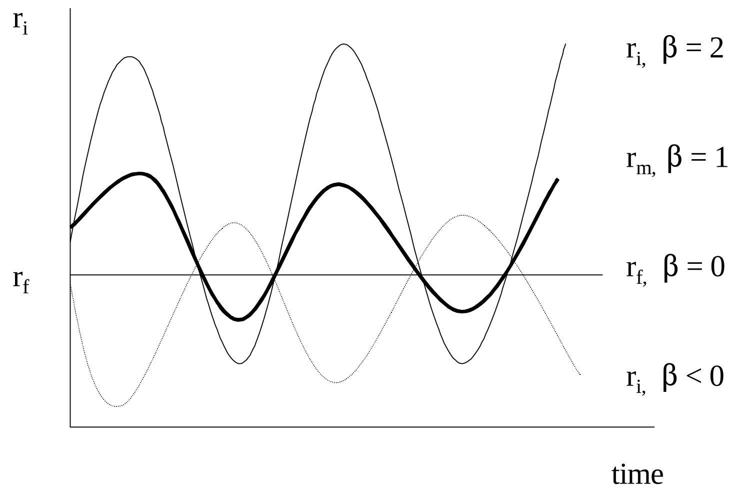

Figure 2
Interpretation of Beta
This figure illustrates the conceptual movement of assets with differing beta over time relative to the movement of the overall market

Again, the first term on the right (equal to $\rho_i^2 \sigma_i^2$ as you may derive by using the definitions of beta and of the correlation coefficient) can be identified as systematic risk, the second term on the right is idiosyncratic risk. Note that the measures of systematic and idiosyncratic risk are slightly different from those in the previous interpretation.

Yet a third way to interpret risk in the CAPM, yielding a similar decomposition, considers the marginal impact of asset *i* in affecting total portfolio risk, as measured by variance. First, using the linearity property of covariance as derived in the Appendix together with the expression of market portfolio return as a weighted average of asset returns, write portfolio variance as:

$$(14) \qquad \sigma_m^2 = \sum_{i=1}^{n} s_i \, \sigma_{im} .$$

Thus,

$$(15) \qquad \partial \sigma_m^2 / \partial s_i = \sigma_{im} = \rho_i \sigma_i \sigma_m .$$

Again similar but not quite identical to the two earlier interpretations of risk. Lastly, note that equation (6.6) below implies that the marginal impact of asset *i* in affecting the *standard deviation* of portfolio risk is given as $\rho_i \sigma_i$, equivalent to our first interpretation.

When is the systematic risk of an asset high? It is easy to check from equation (10) that the beta of a mutual fund representing the market is equal to one (just set *i* = *m* in equation (10)). Thus the "average" asset has a beta of one. Assets with more systematic risk have betas larger than one; assets with less systematic risk have betas less than one. Since covariances can be negative, it is possible for assets to have negative betas (even though we find very few such assets in practice). This occurs when an asset's return tends to move opposite to the market return. Why is that asset's risk premium negative (its expected return will be below the risk free rate)? The reason is that a negative-beta asset can be used to offset some of the risk of other assets within a well-diversified portfolio. Thus, accepting a rate below the risk-free rate is tantamount to buying some insurance. Figure (2) shows, in a naive but illustrative way, how an asset moves with the market depending on its beta.

The graph with $\beta = 2$ indicates an asset without idiosyncratic risk that moves with amplitude of twice that of the market; note that its return, on average, is higher than that of the market. The asset with negative beta moves counter to the market and thus has an average return below the risk free rate. Note that it is necessary that this asset, at least part of the time, should have its return exceed the risk free rate; if not, one could shortsell this asset, borrow at the risk free rate and be guaranteed an arbitrage profit. A graph with $\beta = 0$ (other than the risk free asset, not shown) indicates an asset with only idiosyncratic risk: it is totally out of sync with the market fluctuations, even though it may have higher amplitude, and accordingly has an average return equal to the risk free rate.

A "deeper" explanation of risk in the CAPM context is that, comparing assets with equal mean payoffs, those assets which pay off most when ex-post wealth is highest, are the assets, of course, that co-vary strongly with the market; but high ex-post wealth mean low marginal utility. Thus those assets pay off most when the payoff is least useful (and least when the payoff is most useful). Those assets are considered riskier.

(d) Some Empirical Issues

In empirical work it is standard to use a U.S. stock market index (such as the S&P 500 index, the CRSP value-weighted index, or the CRSP equal-weighted index) as the market portfolio. The CAPM is then tested via a *two-pass* regression method (which will be discussed in more detail later on in this chapter). First, the beta is estimated from a *time series* regression by regressing past asset returns on past market returns, typically using five years of monthly data. The beta is found as the slope coefficient of the regression [as follows from equation (9)]:

$$(16) \qquad r_{it} - r_{ft} = \alpha_i + \beta_i (r_{mt} - r_{ft}) + \varepsilon_{it}.$$

In practice, a similar regression, called the *market model*, is more common:

$$(16') \qquad r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it}.$$

Both regressions based on equations (16) and (16') should yield similar results for the estimates β_i under the assumption that the Sharpe-Lintner CAPM is true and the risk free asset return does not vary over time. In actuality, the risk free return does vary over time but not dramatically so, so that the practical difference between equations (16) and (16') is minimal. While, the market model, equation (16'), is more commonly used by practitioners and academics alike, equation (16) is preferable conceptually as it does not transfer the serially correlated noise due to omission of the risk free rate into the error term of the regression.

In the second stage of the two-pass regression method, formal testing of the CAPM is based on a *cross-sectional* regression, using estimated betas and returns from a cross-section of firms at a given time. These CAPM tests typically use the following formulation. Based on equation (9) and using the estimated betas $\hat{\beta}_i$ from equation (16) as one of the independent variables we can write:

$$(17) \qquad r_{it} - r_{ft} = a_{0t} + a_{1t}\hat{\beta}_i + a_{2t}x_i + \varepsilon_{it} \,,$$

with expected coefficient values $a_{0t} = 0$, $a_{1t} = r_{mt} - r_{ft}$, and $a_{2t} = 0$. The subscripts t are added here because the realized return, market return and risk free rate will generally vary over time. Equation (17) represents the empirical security market line. The following testable implications can be teased out of equation (17): the intercept $a_{0\,t}$ or *alpha* should be zero; the slope of the beta variable a_{1t} should be equal to the realized market excess return $r_{mt} - r_{ft}$; and the slope coefficient a_{2t} of any other explanatory variable should be insignificant. The second implication states that the assets i will all lie on the same empirical security market line. Interestingly, if the realized market excess return $r_{mt} - r_{ft}$ is negative, then higher betas should have lower returns than lower beta securities–such are the workings of risk; the empirical security market line should then have a negative slope. Figure 3 illustrates the empirical SML in the case when the realized excess market return is negative.

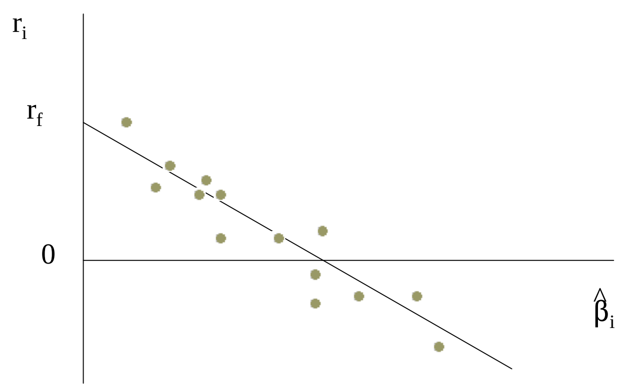

Figure 3
Empirical Security Market Line
Relation between estimated beta of individual securities and their returns. Negative slope expected when the realized market excess return is negative

A quick summary of empirical results is as follows. First, the intercept is often significantly positive but small. Second, the beta slope is often significant but closer to zero than predicted. So, beta does have predictive power but not exactly in the way that the theory suggests: low-beta securities earn more than the CAPM predicts; high-beta securities earn less. Third, the empirical security market line is linear as suggested by the model, meaning that the addition of a beta-squared term (as an x_i variable) is insignificant. Fourth, idiosyncratic risk does not explain return as predicted. A series of other variables, however, does appear to explain returns, in contradiction with the CAPM. Fifth, as shown by Banz (1981) and Reinganum (1981) size affects returns: smaller firms appear to earn higher expected returns than larger firms. Sixth, "value" stocks, with low price-earnings ratios or, similarly, high dividend-price ratios or high book-to-market ratios earn abnormal returns as shown by Basu (1977) and Litzenberger and Ramaswamy (1979), Fama and French (1992). Seventh, Keim (1983) finds that abnormal returns tend to occur in January. In fact, the abnormal returns for small firms occur almost exclusively in the first ten days of January. Eight, Fama and French (1992) found that beta fully loses its predictive power for the recent period in a regression that includes the book-to-market ratio and size variables. Ninth, a recent empirical model by Fama and French (1996) is now popular and is often referred to as Fama and French's "three-factor" model. It includes an asset's market sensitivity (as measured by the standard beta), the sensitivity to excess returns of small firms, and the sensitivity to excess returns of value stocks (high book-to-market stocks) as the three factor affecting an asset's expected excess return.

While heavily contested, the Fama and French results have shown that the CAPM is far from perfect. However, it certainly is not dead (as some have claimed). Sharpe says on this issue: "In the data it's hard to find a strong, statistically significant relationship between measured betas and average returns of individual stocks in a given market. On the other hand it's easy to build a model of a perfectly efficient market in which you could have just that trouble in any period. The noise could hide it." Later in this chapter we will discuss empirical methodology for estimating the CAPM in more detail.

(e) Applications of Beta Estimation and the CAPM

As we shall see in the remainder of this chapter, there are some problems with the testability of the CAPM. This does not mean that the model is not useful. In fact, the CAPM is still one of the most widely applied models in all of economics. The applications of the CAPM can be categorized in the following groups:

1. The Cost of Capital

Capital budgeting is used to tell a firm whether a particular project is profitable. A key variable in any capital budgeting procedure is the cost of capital; or, in economic terms, the opportunity cost of the capital necessary to finance the project. The opportunity cost accounts for time preference as measured by the risk free interest rate and risk. The CAPM implies that relevant risk is systematic risk that can be measured based on the (estimated) beta of the project and the anticipated market excess return.

A related application is in regulation. In a case, for instance, where the government fixes the price of a particular service provided by a utility, the administered price depends on providing the utility with a fair return on capital. This "fair" return is often calculated by applying the CAPM to determine the systematic risk of the utility's activities and thus obtaining the required return.

2. Portfolio Return Evaluation

To determine how a mutual fund or any other managed portfolio perform, it is inappropriate to evaluate realized or average returns of the fund. The reason is that higher levels of systematic risk in the portfolio imply higher average returns. Thus, to evaluate fund performance, a risk correction must be made. Typically, the fund's "alpha" based on the market model is calculated and funds with higher alphas are considered to perform better.

3. Event Studies

Many empirical studies in finance use "event study methodology" [see for instance Fama, Fisher, Jensen, and Roll (1969), Brown and Warner (1980, 1985), and Campbell, Lo, and MacKinlay (1997, Chapter 4)] to determine whether the impact of a particular event is consistent with theory. The basic idea is to verify whether "abnormal" returns are generated in response to the event. In many studies, to account for leakage of information, the cumulative abnormal returns (CARs) over a period stretching from a few days before until a few days after the event are computed; it can then be checked whether the CARs are statistically significantly positive. In these cases the CAPM is not necessary. However, if the event window is substantially more than a few days, excess returns may occur purely due to high beta risk. To adjust for risk and to be able to distinguish abnormal returns from merely excess returns, it is necessary to employ an asset pricing model which is, in practice, usually the CAPM.

(f) Applications and exercises

1. Additivity of beta: Show that the beta of a zero-investment portfolio, holding asset *i* and shorting asset *j* equals the difference of the betas of asset *i* and asset *j*.

2. Extension of earlier question: For the information in question 3.3, find the market portfolio and the beta of risky asset 1.

3. Is it possible for any asset *i* that its portfolio share s_i is negative? Consider this question in: (a) the optimal portfolio choice case discussed in 3(c) or 3(d) of Chapter II; (b) the Sharpe-Lintner CAPM.

2. ALTERNATIVE PROOFS OF THE CAPM

As particular applications often require modifications to the standard model it is useful to look at different proofs so that modifications may be incorporated more easily by adapting the most suitable proof.

(a) A shortcut for the general proof

Start with equation (3.26) in Chapter II.3(d) as representing the CML. Thus every investor will hold a portfolio combining the tangency portfolio and the risk free asset. Consider however a portfolio including in addition an individual asset *i*. The resulting portfolio has the following mean and variance of return:

(1) $$\mu = (1 - s_i - s_T)r_f + s_i\mu_i + s_T\mu_T \,,$$

(2) $$\sigma^2 = s_i^2\sigma_i^2 + 2s_is_T\sigma_{iT} + s_T^2\sigma_T^2 \,.$$

Now raise s_i by reducing s_0 in equations (1) and (2) . This yields

$$(3) \qquad d\mu = (\mu_i - r_f)\, ds_i \ ,$$

$$(4) \qquad d\sigma^2 = 2\sigma d\sigma = 2(s_i\sigma_i^2 + s_T\sigma_{iT})ds_i \ .$$

For an optimal portfolio we know that $s_i = 0$. Thus we can write equations (2) and (4) as:

$$(5) \qquad d\sigma = \frac{s_T\,\sigma_{iT}}{\sigma}\, ds_i\, ; \quad \sigma = s_T\,\sigma_T.$$

Based on equations (5), after eliminating s_T , we find

$$(6) \qquad d\sigma = \frac{\sigma_{iT}}{\sigma_T}\, ds_i \ .$$

Note that in equation (6) $\sigma_{iT}/\sigma_T = \rho_{iT}\sigma_i$ represents the contribution of asset i to portfolio risk, which is only the part of the standard deviation that is correlated with the tangency portfolio.

Combining equations (3) and (6) produces:

$$(7) \qquad \frac{d\mu}{d\sigma} = \frac{(\mu_i - r_f)\,\sigma_T}{\sigma_{iT}} \ .$$

But, from equation (3.26) we also find the slope of the CML as:

$$(8) \qquad \frac{d\mu}{d\sigma} = \frac{\mu_T - r_f}{\sigma_T} \ .$$

Equating the slopes in equations (7) and (8) gives

$$(9) \qquad \mu_i = r_f + \beta_{iT}\,(\mu_T - r_f),$$

with $\beta_{iT} \equiv \sigma_{iT}/\sigma_T^2$. Complete the proof of the CAPM equation by verifying that, since all individuals hold risky assets only in portfolio T, this must be the market portfolio: $T = m$.

(b) A constructive proof when returns are multi-variate normal

We assume here specifically that returns have a multi-variate normal distribution. Thus the proof here is less general than the previous proofs. However, it is more straightforward and self contained. Assume investor k who maximizes expected utility subject to an initial wealth constraint and the requirement that all portfolio shares sum to one:

$$(10) \qquad \underset{\{s_{ik}\}_{i=0}^{n}}{\text{Max}} \; E[u_k(w_k)]$$

$$(11) \qquad \text{s.t.} \quad w_k = \sum_{i=0}^{n} s_{ik}(1+r_i)\,\bar{w}_k$$

$$(12) \qquad \text{s.t.} \quad \sum_{i=0}^{n} s_{ik} = 1$$

After substituting the constraints into equation (10), the first-order conditions for any asset *i* become:

$$(13) \qquad E[u_k'(w_k)(r_i - r_f)] = 0.$$

Using the definition of covariance [see Appendix] we obtain:

$$(14) \qquad E[u_k'(w_k)](\mu_i - r_f) = -\text{Cov}[u_k'(w_k), r_i].$$

Note that equation (14) formalizes the "deep" intuition in Chapter III.1(c). Realizing that the normality of all r_i implies that w_k is also normally distributed, we can apply *Stein's Lemma* [see Appendix]. Thus:

$$(15) \qquad E[u_k'(w_k)](\mu_i - r_f) = -E[u_k''(w_k)]\,\text{Cov}(w_k, r_i).$$

Now define:

$$(16) \qquad \theta_k \equiv -E[u_k''(w_k)]/E[u_k'(w_k)].$$

This term is similar (but not equal due to the expectations that are taken) to the coefficient of absolute risk aversion. Using equation (16) in equation (15) yields:

$$(17) \qquad \sum_{k=1}^{K} \theta_k^{-1}(\mu_i - r_f) = \text{Cov}(w_m, r_i) = \bar{w}_m\,\text{Cov}(r_m, r_i),$$

which follows since $w_m = \Sigma_k w_k = \bar{w}_m(1 + r_m)$. When we use equation (17) for asset *m* we get

$$(18) \qquad \mu_m - r_f = [\sum_{k=1}^{K} \theta_k^{-1}]^{-1}\,\bar{w}_m\,\sigma_m^2.$$

Note that equation (18) provides an aggregate measure of risk aversion that we did not encounter in the previous proof. [In particular, if we were to assume CARA preferences, so that θ_k would become a constant, then we would have an explicit expression for aggregate risk aversion and explain the aggregate market risk premium as the product of aggregate risk (market risk) and aggregate risk aversion]. Dividing equation (17) by equation (18) to eliminate the θ_k terms produces:

(19) $\mu_i = r_f + \beta_i (\mu_m - r_f)$,

with the standard definition of beta.

(c) A Quick General Equilibrium Version of the Basic Proof

Start with equation (II.3.30), pertaining to the efficient portfolio demands of an individual investor k :

(20) $\boldsymbol{s_k}^{\boldsymbol{T}} \Sigma = \lambda_k \boldsymbol{e}^{\boldsymbol{T}}$.

Aggregate over all individuals to obtain:

(21) $\boldsymbol{a}^{\boldsymbol{T}} \Sigma = \Lambda \boldsymbol{e}^{\boldsymbol{T}}$,

where $\Lambda = \sum_{k=1}^{K} \lambda_k$ and $\boldsymbol{a} = \sum_{k=1}^{K} \boldsymbol{s_k}$. Since $\boldsymbol{a}$ is a vector of the aggregate value of each risky asset in equilibrium, the left-hand-side of equation (21) represents the appropriately weighted row vector of covariances of the risky asset returns with the market return (see Appendix for the appropriate covariance definition), with typical element σ_{im} . Thus, equation (21) for a typical row vector element is

(22) $\sigma_{im} = \lambda e_i$.

Post-multiplying equation (21) by $\boldsymbol{a}$ yields:

(23) $\boldsymbol{a}^{\boldsymbol{T}} \Sigma \boldsymbol{a} = \Lambda \boldsymbol{e}^{\boldsymbol{T}} \boldsymbol{a}$,

which is equivalent to:

(24) $\sigma_m^2 = \lambda e_m$.

Dividing both sides of equation (22) by the same sides of equation (24) produces the desired CAPM equation.

(d) Applications and exercises

1. Prove the CAPM for quadratic preferences using the set-up in equations (10) - (12). Do not assume here that returns are normally distributed.

2. For the proof in section 2(b), state where assumptions (1) - (13) are introduced.

3. THE ZERO-BETA CAPM

One of the debatable assumptions of the CAPM is the hypothesis that a risk-free asset exists. In spite of the existence of, say U.S. T-Bills with any desired short maturity, one could easily argue that no truly risk free asset exists. First, there is inflation risk. One might of course hold an indexed security but available maturities for such securities are limited and inflation corrections may not be appropriate for the individual investor. For instance, the overall CPI may not be very relevant for a retiree living in Alaska. Second, there is reinvestment risk. A short maturity is not riskless for someone saving for retirement as the available interest rate upon maturity is not known. On the other hand, a longer maturity is risky if there is a chance that liquidity is needed ahead of retirement, since selling a long-term bond before maturity may involve a substantial capital loss. Third, the issuer, say the U.S. government, may default in the case of a major natural disaster or war. In addition, the "risk free" rate and the market return may not even be independent. Inflation, for instance, might affect both rates in the same direction. We thus drop assumption 10 of the basic CAPM and examine the resulting asset pricing model. This was first accomplished by Black (1972) and the resulting model is called the "zero-beta" CAPM (or the Black CAPM, as opposed to the regular CAPM which is usually referred to as the Sharpe-Lintner CAPM) to reflect the fact that, in this model, the role of the risk free asset is taken by a portfolio that is uncorrelated with the market and which thus has zero beta.

(a) Derivation

Consider any frontier portfolio as discussed in section 3(c) of Chapter II. The covariance between the return on an asset *i* and the frontier-portfolio return is given as:

$$(1) \qquad \sigma_{ip} = \boldsymbol{s_i}^{\boldsymbol{T}} \Sigma \, \boldsymbol{s_p}^{*}.$$

Using the transpose of equation II.3.18 we can then write:

$$(2) \qquad \sigma_{ip} = \lambda \, \boldsymbol{s_i}^{\boldsymbol{T}} \boldsymbol{\mu} + \kappa \, \boldsymbol{s_i}^{\boldsymbol{T}} \mathbf{1},$$

which becomes, using the definition of portfolio return and the fact that all portfolio shares add to one,

$$(3) \qquad \sigma_{ip} = \lambda \, \mu_i + \kappa.$$

If we let asset *i* be the frontier portfolio *p* itself, then equation (3) implies:

$$(4) \qquad \sigma_p^2 = \lambda \, \mu_p + \kappa.$$

Now define any portfolio *z* that is uncorrelated with the frontier portfolio *p*. Again using equation (3) gives:

$$(5) \qquad \sigma_{zp} \equiv 0 = \lambda \, \mu_z + \kappa \quad \rightarrow \quad \mu_z = -(\kappa/\lambda).$$

Divide equation (4) by equation (3) to obtain "beta"; then divide numerator and denominator by κ and employ equation (5) to eliminate the κ/λ terms. This produces:

$$(6) \qquad \mu_i = \beta_{ip}\mu_p + (1-\beta_{ip})\mu_z, \quad \beta_{ip} \equiv \sigma_{ip}/\sigma_p^2 .$$

We can write in linear regression format:

$$(7) \qquad r_i = b_0 + b_1 r_p + b_2 r_z + e_i .$$

Here, we have the following testable implications. Since r_p and r_z are uncorrelated we know that $b_1 = \beta_{ip}$ and $b_2 = \beta_{iz}$. And $b_1 = 1 - b_2$. Further, since the expected regression residual is zero by definition, equation (6) implies that $b_0 = 0$. Note also that we now know from the regression properties that it is possible to write:

$$(8) \qquad r_i = \beta_{ip} r_p + (1-\beta_{ip}) r_z + \varepsilon_i ,$$

$$\beta_{ip} \equiv \sigma_{ip}/\sigma_p^2, \quad E(\varepsilon_i) = E(\varepsilon_i r_p) = E(\varepsilon_i r_z) = 0 .$$

The zero-covariance portfolio is of obvious importance here. It may be found graphically as follows. Assume that *p* is an efficient portfolio. Then, in mean-standard deviation space, draw the line tangent to *p*. The intercept of the tangent line would be the analogy of the risk free rate if a risk free asset existed and if *p* were the market portfolio. Then the expected return on a portfolio *z* would be found by extending a horizontal line from the "risk free rate" μ_z to any

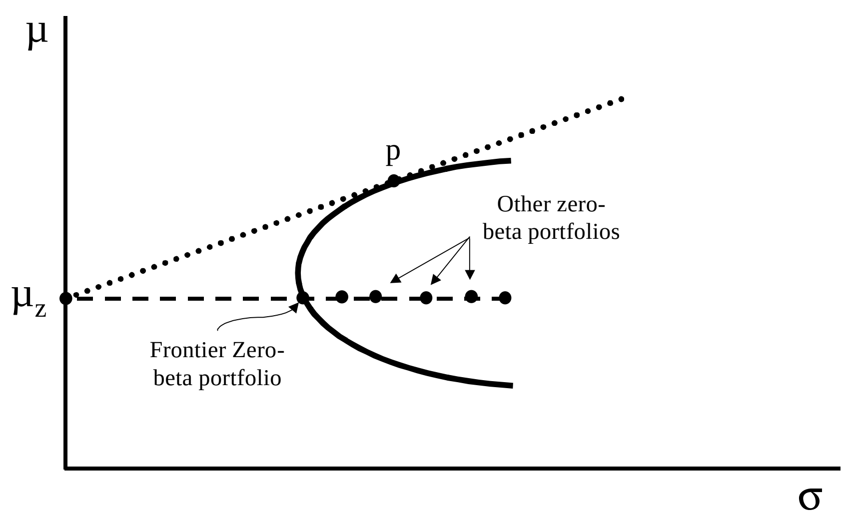

Figure 4
Zero-Covariance Portfolio
Assuming portfolio p is an efficient portfolio, the intercept of its tangent line is analogous to the risk-free rate.

feasible portfolio on or inside the portfolio frontier, as shown in Figure 4. The intuition is that a portfolio return having zero covariance with the market should have the same expected return as the risk free asset. It can be shown that this graphical result is true even if p is not the market portfolio. It can also be shown that a unique *frontier z* can be found (that is, a z that is exactly on the portfolio frontier) for given p and that, if p is efficient, then z is inefficient, and vice versa. These two results can be proven as an *exercise* as formulated below.

As in the basic CAPM we can take p to be the market portfolio. The reason is that the market portfolio must be on the portfolio frontier. Define *market wealth* as the aggregate level of wealth:

$$(9) \qquad \bar{w}_m \equiv \sum_{k=1}^{K} \bar{w}_k \,,$$

where the individual initial wealth of the K investors in the economy is summed to get (initial) market wealth. Consider next the aggregate quantity of any asset i held in equilibrium. This is given as the equilibrium market share of asset i, s_{im}, times market wealth. Since in equilibrium *all assets are held* it must be true for any asset that:

$$(10) \qquad \sum_{k=1}^{K} s_{ik}\bar{w}_k = s_{im}\bar{w}_m \quad \leftrightarrow \quad \sum_{k=1}^{K} s_{ik}(\bar{w}_k/\bar{w}_m) = s_{im} \,.$$

Thus, the portfolio consisting of the market shares of all assets (the *market portfolio*) is a convex combination of the portfolios of all individuals [convex since $\Sigma_k(\bar{w}_k/\bar{w}_m) = 1$ from equation (9)]. We know that, in equilibrium, and using homogeneity assumptions 11 and 12 of the CAPM, all individuals hold frontier portfolios. (In fact, since the whole frontier can be traced out by the various linear combinations of holdings of only two different frontier portfolios, we have a two-fund separation result which explains that asset pricing is determined by two factors only). Since a convex

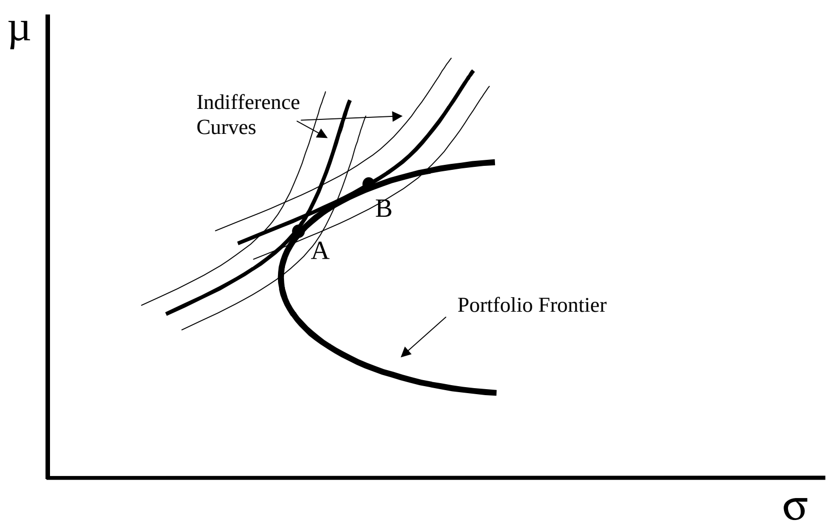

Figure 5
Portfolio Choice and the Market Portfolio
Optimally chosen portfolios must lie on the efficient frontier; the convex combination of such portfolios must lie on the efficient frontier as well. Thus, the market portfolio, being one such convex combination, is an efficient portfolio.

combination of a frontier portfolio is still a frontier portfolio we now know that the market portfolio is a frontier portfolio.

Figure 5 illustrates the optimal portfolio choices of two arbitrary individuals. Both face the same opportunity set of risky assets but may have different mean-variance preferences. Both choose a point on the efficient frontier; as the combination of two (efficient) frontier portfolios is still an (efficient) frontier portfolio [see Chapter II, section 3(b)], the portfolio of their pooled assets is also on the (efficient) frontier. Adding one by one the assets of all other individual in this manner produces the market portfolio that must thus be on the (efficient) frontier. Thus we can replace equation (6) by

$$(11) \qquad \mu_i = \beta_i \mu_m + (1 - \beta_i)\mu_{zm}, \qquad \beta_i \equiv \sigma_{im}/\sigma_m^2 ,$$

with some obvious changes in notation.

(b) Empirical Implementation

Empirically, one may obtain portfolio *z* by constructing a portfolio frontier, then taking any frontier portfolio *p* and finding the unique frontier asset uncorrelated with it. Here *p* does not have to be the market portfolio although it is often convenient to use the market portfolio as such. More often, though, in empirical implementation, the zero-beta portfolio is omitted in the market model regression. To see why this is possible, consider again equation (8), re-stated here for the frontier portfolio being the market portfolio

$$(12) \qquad r_i = \beta_i r_m + (1 - \beta_i) r_z + \varepsilon_i \ .$$

By construction of the zero-covariance portfolio, $\sigma_{zm} \equiv Cov(r_z, r_m) = 0$. As a result, the regression coefficient on r_m is not subject to an omitted variables bias if r_z is omitted from the regression. Thus, in principle, any security's beta can be estimated without bias from the following market model regression:

$$(13) \qquad r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it} .$$

The variables used in this regression to obtain the Black CAPM betas must be the *real* market and security returns. In contrast, the variables for obtaining the Sharpe-Lintner CAPM betas are the market and security *excess* returns in equation (1.16).

(c) Applications and exercises

1. Derive mathematically that the method for finding the zero-beta portfolio graphically as in Figure 4 is correct for any frontier portfolio *p*. [Provide Hints].

2. Show that *z* is unique for given *p* and that, if *p* is efficient, then *z* is inefficient, and vice versa. [Provide Hints].

3. Provide alternative proofs for equation (13) using: (a) the method in section 2(a) [see Copeland and Weston (1992)] and (b) the method of section 2(b).

4. OTHER ISSUES IN THE BASIC CAPM

(a) The Roll Critique

The shallow version of the Roll Critique is that the CAPM is not testable because the proxy used for the market return is imprecise. The market portfolio should consist of all risky assets, including bonds, precious metals, real estate, human capital, and international stocks. Instead the market proxy used in practice only includes U.S. stocks, often limited to those traded on the major exchanges only (NYSE and, more recently, AMEX and NASDAQ). An imperfect approximation of the market portfolio leads to an imperfect measure of market return, especially if returns are equal-weighted rather than value-weighted. Clearly, with an inadequate market return proxy, and the impossibility of getting a much better one, any apparent rejection of the CAPM could be defended by saying that results are biased due to measurement error related to the improper measure of the market return.

Roll's true critique, however, is more extensive. One result of the mathematics of the portfolio frontier (which we haven't proven here) is that there is a positive linear relation between any two different assets or portfolios based on the beta between these two assets, where only one of the two assets needs to be an efficient portfolio. This relation is tautological and would be true even if, say, all individuals were risk neutral! Thus, assets might line up on the SML even if the CAPM is not true. This would not happen, however, if the benchmark asset were not an efficient portfolio. Roll suggests that the (only) way to test the CAPM is to check if the market portfolio is efficient. But if the portfolios constructed in the traditional CAPM tests are efficient (which, likely, they would be as they are chosen as an equal-weighted average of a large group of portfolio), then seeing if the portfolios line up on the market line would be misleading. Thus, the efficiency of the CAPM would have to be examined directly (by trying to find a portfolio with lower variance given the mean return of the market portfolio). But here is where the imperfect nature of the market proxy is particularly damaging.

Figure 6 illustrates the dual problem in testing the CAPM: the tautological nature of the linear CAPM relation together with the problem of measuring market return exactly. In the case of the zero-beta CAPM for instance, if the CAPM is true, equation (3.12) should hold:

$$(1) \qquad r_i = r_z + \beta_i (r_m - r_z) + \varepsilon_i .$$

Equation (1) should hold in the sense return depends linearly on β_i and that any other explanatory variable outside of β_i can have no impact. Roll's analysis shows however that *equation (1) will hold identically if and only if the market portfolio is mean-variance efficient.* Given that we are using actual data to estimate means and covariances, equation (1) will hold identically if and only if the market portfolio is *ex-post* mean-variance efficient; that is, given that the market portfolio ends up on the efficient frontier (even if just by coincidence), equation (1) should give an R-squared of 1. As Figure 6 then summarizes, equation (1) should always fail (even if the CAPM is true) if the market proxy is not ex-post efficient; and should always hold (even if the CAPM is false) if the market proxy is ex-post efficient.

	Market Proxy Ex-Post Efficient	Market Proxy **Not** Ex-Post Efficient
CAPM **True**	**(1) Holds**	**(1) Fails**
CAPM **False**	**(1) Holds**	**(1) Fails**

Figure 6
The Roll Critique
The result of testing the CAPM does not depend on whether or not the CAPM is true. It depends only on whether or not the market proxy is ex-post mean-variance efficient

Thus, failure of an equation like (1) to hold tells us nothing about whether the CAPM is true or false, and neither does the event of an equation like (1) holding tell us anything about whether the CAPM is true of false.

The approach of Fama and MacBeth (1973) discussed in the upcoming section may provide a true test of the *usefulness* of the CAPM as it relies on lagged measures of betas to forecast future expected returns. Clearly such approach is not tautological. If betas are stable over time then returns can be forecast based on beta. One problem with this, however, is that beta may just proxy for higher expected return. Say that the CAPM is false and that firm size is the only real determinant of return (smaller firms are, somehow, riskier). Then smaller firms would have higher returns, but, because of the mathematics of the portfolio frontier, would also have higher betas. If we use these betas in the future we would likely still find higher returns for higher-beta firms. Not because of the CAPM but because "high-beta" firms were small in the previous period and still will be small in the next period and thus tend to have higher returns.

The previous argument, while it makes testing the CAPM more difficult, is not fatal. Certainly if the realized market risk were always positive it would be impossible to separate the firm size effect from the beta effect. However, if the market risk premium realization is negative, high beta assets should have lower returns while small firms should have higher returns just as before. Thus, more generally, we know from the mathematics of the portfolio frontier that high measured beta may proxy for high past returns. If these high returns have a systematic cause, then beta proxies for this cause. But, unless this cause is somehow strongly correlated with the realized market return, a negative market return realization should cause high-beta assets to have lower returns whereas no such relation is likely under the alternative.

(b) Applications and exercises

1. Concerning the following statements, explain whether they are *True*, *False*, or *Ambiguous.*

   (a) In the Sharpe-Lintner CAPM, if the risk free rate is lowered, all else equal, stock returns on average will be unchanged.

   (b) Returns are not elliptically distributed and preferences are not quadratic, but the market portfolio is mean-variance efficient. Then the linear CAPM equation will hold.

   (c) When a risk free asset exists, there exists no portfolio (other than the risk free asset itself) which has zero beta with the market portfolio.

   (d) Empirical estimation of the Black version of the CAPM requires that first a zero-beta portfolio must be identified.

   (e) In the Sharpe-Lintner CAPM, all negative beta securities (if any exist) will underperform the market portfolio at every point in time.

   (f) Suppose there is a systematic source of risk (such as inflation risk) that is uncorrelated with market risk. In this case the CAPM can still be true.

5. EMPIRICAL METHODOLOGY IN ESTIMATING THE CAPM

The standard methodology in estimating the CAPM or one of its extensions is the *two-pass regression method*. One may criticize this method, but fact is that it is currently the standard in the finance literature. This empirical approach was developed by Black, Jensen, and Scholes (1972) and refined by Fama and MacBeth (1973). The Fama-MacBeth approach, suitably adapted, is the method of choice in empirical asset pricing; any deviations from this methodology should be well motivated.

It must be understood that, based on the Roll Critique, a reliable *test* of the CAPM is not a possibility. Nevertheless, the empirical approach outlined here provides a numerical evaluation of the usefulness of a particular CAPM formulation. I will say more about this at the end of the section. For now, we must keep in mind that the empirical content of the Sharpe-Lintner version of the CAPM is the following:

$$(1) \qquad r_{it} - r_{ft} = a_{0t} + a_{1t}\beta_{it} + a_{2t}x_{it} + e_{it}\,, \quad a_{0t} = 0\,,\ a_{1t} = r_{mt} - r_{ft}\,,\ a_{2t} = 0\,.$$

As Roll (1976) has shown, equation (1) will hold empirically, if and only if the proxy chosen to represent the market portfolio is on the portfolio frontier. Whether the proxy is on the frontier is evaluated ex post, meaning that a frontier is constructed from the actual observations used in testing the model and that the location of the realized mean and standard deviation of the proxy return relative to the frontier is considered. Clearly, if the proxy is close to efficient, equation (1) will hold approximately; and, if the proxy is not close to being efficient, equation (1) will not hold even approximately.

The two-pass regression methodology focuses on the "testable" implications in equation (1). In the first "pass" time series estimates $\hat{\beta}_{it}$ of individual asset betas (and, if necessary, of the $\hat{x}_{it}$) are obtained; in the second "pass" these beta estimates are employed in a cross-sectional regression to obtain parameter estimates $\hat{a}_{0t}$, $\hat{a}_{1t}$, $\hat{a}_{2t}$ which are averaged over time, yielding $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$. These parameter estimate averages are finally compared statistically to their predicted values of 0, $\bar{r}_m - \bar{r}_f$, and 0, respectively. This method in its simplest version is employed in Mankiw and Shapiro (1986). They ignore, however, a series of thorny empirical issues that is handled more appropriately by Fama and MacBeth (1973).

I next outline step by step the gist of Fama and MacBeth's (1973) approach and indicate where it deviates materially from the Black, Jensen, and Scholes (1972) version. I also attempt to articulate the empirical rationale for the various complexities introduced. The whole method is presented in a "cook book" way so that it can be applied straightforwardly in a variety of different CAPM applications. Figure 7 provides a time line to summarize the different steps employed in the Fama-MacBeth approach to be discussed next.

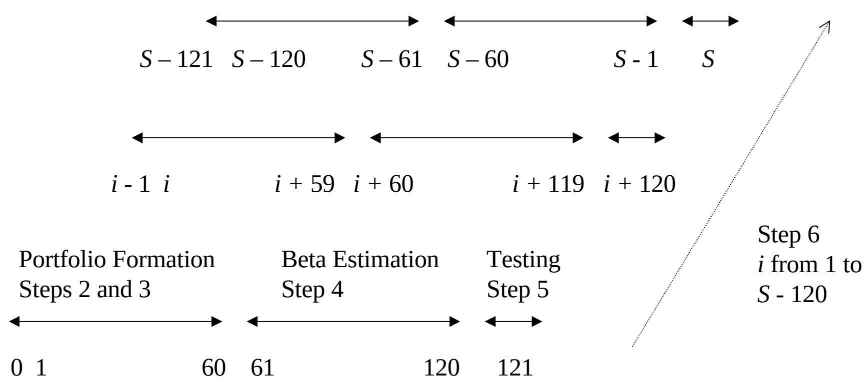

Figure 7

The Fama-MacBeth Methodology

The time line indicates the different steps in testing the CAPM. The output from Step 5 is obtained repeatedly for all *i* from 1 to *S* – 120 to produce a distribution of regression coefficients that can be evaluated statistically

Step 1. Data

Obtain *all data* suitable for the purpose at hand. Data required are time series of the following: returns for various assets, a market proxy (typically a stock index such as the S&500 or the CRSP value-weighted index; the early empirical work employed the CRSP equal-weighted index), a risk free rate for the Sharpe-Lintner CAPM (ideally the return on a riskless bond with exactly a month left to maturity; in practice, the return on 3-month T-Bills or 1-month T-Bills or, in periods where these returns are not available, dealer commercial paper rates) or an inflation rate in the case of the zero-beta CAPM, and (if appropriate) interesting x variables. Typically, for all but the x variables, these data can be pulled from CRSP. There are two reasons for employing as many data as are reasonably available. First, to increase the power or accuracy of the statistical results; Second, to avoid suspicion of "data mining" in the form of restricting your sample to the subset that gives the best results.

In selecting the data, it is necessary to determine in advance how long of a time series is required for a particular security to be included in the estimation. In CRSP the number of securities listed at a particular point in time varies. New securities are listed; others are delisted due to merger, bankruptcy, or exchange-specific rules for listing. It is important to avoid selection biases, such as a "survivorship" bias that arises, for instance, if only securities are used that were listed continuously from 1926 until now: returns for these securities are biased upward as they have been successful for a long period. If we are prevented from considering only securities that were continuously listed over the whole sample, securities must be included which have missing return observations for part of the sample period. In these cases (which includes most, since, of several thousand securities currently listed on CRSP, only some thirty have been listed continuously since 1926), there must be a clear criterion for when the security should be included. As an example Fama and MacBeth include any security listed at the time when the second "pass" regression is run, if it also has at least 84 earlier data points (60 data points prior, for estimation of its beta, plus an additional 24 data points before that, for portfolio sorting purposes)

Step 2. Preliminary beta estimation

Estimate betas for each asset at each point in time, using time series data in a version of the "market model":

$$\text{(2a)} \qquad r_{it} - r_{ft} = \alpha_{iT} + \beta_{iT}(r_{mt} - r_{ft}) + \varepsilon_{it},$$

for the Sharpe-Lintner version of the CAPM. Note that T is the final data point in the sub sample used to estimate a beta. Typically, 60 monthly time series observations are used to estimate a beta. Thus the first usable beta in the sample would be $\hat{\beta}_{i\,60}$ estimated from sample points 1 through 60. To obtain $\hat{\beta}_{i\,61}$, we roll the sample forward by one period, using sample points 2 through 61. The reason for using 60 sample points to estimate beta and not increasing this number as we roll the sample forward is that betas are presumed to change over time (referred to as the "nonstationarity" of beta). The choice of 60 sample points reflects the tradeoff between estimation efficiency, for which a longer series may be better, and beta nonstationarity, for which a shorter series may be better. A Weighted Least Squares (WLS) regression, in which less weight is put on sample returns that are further back, seems to present a more efficient approach in the face of nonstationarity in betas. However, such an approach is not common, possibly because establishing the proper weights is a nontrivial matter.

The returns in equation (2a) are all nominal. An inflation correction is not necessary since the excess returns (both the excess return for asset *i* and the market excess return) employed in the equation are automatically in real terms. Whether the *x*-variable in equation (1) needs to be included in equation (2a) is a matter of what the true null hypothesis is. For the Sharpe-Lintner CAPM, no *x*-variable enters under the null that the CAPM is true (in a multi-beta variant of the CAPM that would of course be different), so that no *x*-variable appears in equation (2a). However, under the null that the CAPM is true, we could also set $\alpha_{iT} = 0$. While this may cause the estimates $\hat{\beta}_{iT}$ to be more efficient if the null is exactly true, the inconsistency in beta estimation arising from restricting the constant to be zero in a situation where that is not exactly appropriate, are too severe. Hence, in practice, we estimate (2a) with constant and one slope coefficient as indicated.

If we estimate the Black zero-beta CAPM instead of the Sharpe-Lintner version, equation (2a) becomes:

$$(2b) \qquad r_{it} - \pi_t = \alpha^z_{iT} + \beta^z_{iT}(r_{mt} - \pi_t) + \varepsilon^z_{it} \quad ,$$

where π_t indicates the inflation rate over the period for which the return is measured. Note that here $\alpha^z_{iT} = (1 - \beta^z_{iT})(r_{zt} - \pi_t)$, with r_{zt} representing the return on an asset that has zero correlation with the market. Leaving out the variable $r_{zt} - \pi_t$ in this case presents no "omitted variables" bias since the omitted variable is uncorrelated with the included right-hand side variable $r_{mt} - \pi_t$. And so, again, in practice, we estimate (2b) with constant and one slope coefficient as indicated.

In this step we thus obtain, in principle, $\sum_{t=60}^{S-1} I_t$ different beta estimates, where I_t indicates the number of different assets in the sample at time *t*, and *S* indicates the number of time periods in the sample; the 1 is subtracted as no beta needs to be estimated at the very end of the sample. In practice, for reasons of programming or data management convenience, we sometimes may want to write programs that do not estimate all of these betas before the other steps are completed.

Step 3. Portfolio sorting

Rank assets by beta from high to low. Then split all assets into a given number *P* of portfolios, usually 10 [as in Black, Jensen, and Scholes (1972)] or 20 [as in Fama and MacBeth (1973)]; where portfolio 1 includes the $N_t \equiv I_t/P$ assets with the highest betas (in period *t*) and so on down until portfolio *P* is formed from the N_t assets with the lowest betas. Note that it is assumed that $N_t \equiv I_t/P$ is an integer (that is, there is no remainder); if not, the remaining assets can be allocated such that the some of the beta portfolios have one more asset.

The rationale for forming portfolios is to reduce measurement error in the betas. In equation (1), we need a measure of the beta of each asset for a given time period; but this beta is estimated from a time series regression. Thus we have:

$$(3) \qquad \beta_{it} = \hat{\beta}_{it} + \eta_{it} \quad .$$

As a result, the coefficient a_1 will be estimated inconsistently. To see this consider that, with equation (3), using $\hat{\beta}_{iT}$ to estimate equation (1) implies a theoretical error of $a_1^2\,\sigma_\eta^2 + \sigma_e^2$. OLS, in choosing $\hat{a}_1$, will select it in part to reduce the $a_1^2\,\sigma_\eta^2$ term; thus biasing the estimate toward zero. To minimize this measurement error problem, equal-weighted portfolios are formed so that betas with substantially less measurement error can be calculated for these portfolios, since the idiosyncratic measurement errors are averaged out over a large group of N_t assets. The reduction in measurement error works if the "signal-to-noise" ratio of the betas improves. If portfolios are picked at random, the measurement error may not be reduced since the betas will typically average to around one in large portfolios and so we end up with 10 or 20 portfolios, all with betas around one. Clearly, then, while the noise in beta estimation is reduced, the signal in beta estimation is reduced as well. To maximize the signal in beta estimation it is therefore important to select portfolios to maximize variation in the portfolio betas; this is accomplished by first ranking assets by beta before constructing portfolios from the ranked assets, as is done in this step. A formal analysis of these issues is provided in the appendix of Black, Jensen, and Scholes (1972).

Step 4. Estimating portfolio betas

This step is necessary to complete the first "pass" of the two-pass regression method. A portfolio's beta can be calculated directly as the average of the betas of its component assets (see exercise 1.1 in Chapter III). However, doing this would cause another measurement error issue: the assets with the most extreme beta estimates are most likely to have substantial measurement error; thus in forming portfolios we are systematically grouping assets with, currently, similar beta measurement errors. As a result, a "regression to the mean" problem arises: betas in high-beta portfolios tend to be over-estimated and betas in low-beta portfolios tend to be under-estimated; hence, slope estimates of the impact of beta will be biased downward. The solution to this problem is to estimate portfolio betas with new data.[3] That is, with data outside the sample in which the portfolios where selected:

(4) $$r_{it} - r_{ft} = \alpha_{iT} + \beta_{iT}(r_{mt} - r_{ft}) + \varepsilon_{it}, \quad \text{for all } T \text{ with } t \in \{T\text{-}59, T\}, \text{ and } T \geq 120,$$

(5) $$\hat{\beta}_{pt} = \frac{\sum_{i = pN_t + 1}^{(p+1)N_t} \hat{\beta}_{it}}{N_t}, \quad \text{for all } p \in \{1, P\}.$$

By estimating an equation such as (4), both Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) avoid the second measurement error problem. They ignore the data points used in estimating the preliminary betas for portfolio sorting and, instead, re-estimate portfolio betas with a later part of the time series. Black, Jensen, and Scholes use all data points beyond the 60 used in estimating the preliminary betas; Fama and MacBeth use 60 data points beyond the 60 used in obtaining portfolio betas, as equation (4) indicates. I will discuss the Fama and MacBeth version from here on.

Use $\hat{\beta}_{i\,60}$ from sample points 1 through 60 to sort all assets into P portfolios. The portfolio betas then are obtained as the straight average of the individual asset betas in the portfolio over periods 61 through 120. Thus we obtain $\hat{\beta}_{p\,120}$ as the first "usable" beta for each portfolio. The next beta for each portfolio is obtained by first sorting based on $\hat{\beta}_{i\,61}$ and then estimating $\hat{\beta}_{p\,121}$. Accordingly, in step 4, $P \cdot (S - 119 - 1)$ different beta estimates are generated, where P indicates the number of different portfolios, S indicates the number of time periods in the sample,

[3] Litzenberger and Ramaswami (1979) and Shanken (1992) provide an alternative approach to dealing with the measurement problem by adjusting the standard errors for the bias arising from measurement error.

and 1 indicates the time series point at the very end of the sample for which no beta estimate is needed in the Fama-MacBeth approach.

The exact procedure in Fama and MacBeth (1973) is a little different in the details. One reason is that they were concerned with reducing computation time; an issue which is less of a concern nowadays. For instance, Fama and MacBeth only re-sorted portfolios once every four years, whereas the above procedure implies re-sorting portfolios every month. While the above procedure takes extra computing time, it is more powerful and easier to program.

Step 5. Cross-sectional regressions.

In the second "pass" of the two-pass regression method, the beta estimates are employed as independent variables to explain the cross-sectional variation in the returns of the constructed portfolios. Here the Black-Jensen-Scholes and Fama-MacBeth approaches differ most clearly. Black, Jensen, and Scholes perform the cross-sectional regressions over all time periods used in estimating the portfolio betas (but still exclude the first 60 data points reserved for estimation of the pre-sorting betas). Fama and MacBeth do not use any of the time periods, reserved for sorting portfolios and estimating an associated set of portfolio betas, in the cross-sectional regressions that employ this set of portfolio betas. I here discuss their approach. For each remaining period, coefficient estimates $\hat{a}_{0t}$, $\hat{a}_{1t}$, $\hat{a}_{2t}$ are estimated based on:

$$(6) \qquad r_{pt} - r_{ft} = a_{0t} + a_{1t}\hat{\beta}_{pt} + a_{2t}x_{pt} + e_{pt}\,, \quad \text{for all } t \geq 120.$$

Note that the $\hat{\beta}_{pt}$ variable is obtained from lagged information only, that is using time period $s < t$ only. For an x variable Fama and MacBeth utilize the standard deviation of the estimated residual from the regression based on equation (4), averaged over all assets in the portfolio; this variable describes non-systematic risk. They also use the squared values of the $\hat{\beta}_{it}$ averaged over all assets in a portfolio as an additional x variable to represent possible non-linearities in the empirical asset pricing equation.

The reasons for estimating the cross-sectional equation out of the sample used for generating the beta estimates are not very clear. Fama and MacBeth argue that it is not necessary for "positive" reasons (i.e., finding out to which extent the model is useful in describing actual return data) to predict equation (6) out of sample. For "normative" reasons (i.e., employing the model as an instrument in making better decisions), however, "... the model only has content if there is some relationship between future returns and estimates of risk that can be made on the basis of current information." [Fama and MacBeth (1973, p.618]. Another argument for predicting out of sample may be to avoid possible ways in which in-sample beta estimates somehow "contaminate" the slope estimates in the cross-sectional regression. Finally, the out-of-sample approach avoids some of the sting of the Roll (1976) critique. Even if returns are picked purely randomly, a relation like equation (1) holds ex post if the market proxy turns out to be efficient. Obviously, then, if future returns are also picked randomly, betas estimated from previous data have no predictive value. If betas do have predictive value in the Fama-MacBeth approach, there must be value to the CAPM. An alternative way of expressing this idea is by pointing out that, in the Fama-MacBeth approach, two hypotheses must hold true for equation (1) to pass the empirical test: the market proxy has to be efficient and market betas must be relatively stable. The latter implies that the CAPM is valuable even if it is mostly tautological. There is one drawback to the out-of-sample approach, namely that measurement error caused by instability of the betas over time is increased. This issue seems to be ignored in the current literature.

Estimation of equation (6) yields a set of coefficient estimates $\hat{a}_{0t}, \hat{a}_{1t}, \hat{a}_{2t}$ for all $t \geq 121$. So that we end up with *three* (or more if there are more than one *x* variable) times $S - 120$ coefficient estimates. Notice that $\hat{a}_{1t}$ is not expected to be positive in any period in which the market excess return was non-positive. In fact, at any time when the realized market excess return is negative, $\hat{a}_{1t}$ should be negative as well.

Step 6. Averaging cross-sectional regression coefficients.
The three or more coefficients are now averaged over all $S - 120$ time periods to provide the most powerful test of the CAPM, that is, equation (1). This yields $\hat{a}_0, \hat{a}_1, \hat{a}_2$. The CAPM is rejected if $\hat{a}_0$ deviates significantly from zero; if $\hat{a}_2$ deviates significantly from zero; or if $\hat{a}_1$ deviates significantly from $\bar{r}_m - \bar{r}_f$. Significance here is based on the *t*-statistic for the null hypothesis:

$$(7) \qquad t(\hat{a}_0) = \frac{\hat{a}_0}{\hat{\sigma}_{a_0}/\sqrt{S-120}} \quad , \quad t(\hat{a}_1) = \frac{\hat{a}_1 - (\bar{r}_m - \bar{r}_f)}{\hat{\sigma}_{a}/\sqrt{S-120}} \quad , \quad t(\hat{a}_0) = \frac{\hat{a}_2}{\hat{\sigma}_{a_2}/\sqrt{S-120}} \, ,$$

$$(8) \qquad \hat{\sigma}^2_{a_i} = \frac{\sum_{t=1}^{S-120} (\hat{a}_{it} - \hat{a}_i)^2}{(S-120)(S-119)} \qquad \text{for all } i.$$

As Fama and MacBeth point out, these *t*-statistics provide a bias toward rejecting the model, given that in fact empirical return distributions are not normal but have thick tails (measurement error in the beta estimates has the opposite effect).

If any of the null hypotheses is rejected, the CAPM should formally be rejected as well. This implies then either that the market proxy is not efficient, or that betas are not stable enough to be useful in forecasting. While Black, Jensen, and Scholes formally reject the CAPM, concluding that their $\hat{a}_0$ significantly exceeds zero and their $\hat{a}_1$ is significantly less than $\bar{r}_m - \bar{r}_f$, Fama and MacBeth find no significant deviations for $\hat{a}_0$ and $\hat{a}_1$ and, moreover find that $\hat{a}_2$ does not deviate significantly from zero– both the measure of non-systematic risk and the average square of the estimated betas (as a variable measuring possible nonlinearity) are insignificant. Thus, market betas appear to be useful in determining expected returns on the sort of assets, U.S. stocks traded on the NYSE, considered in the sample. A rejection, by the way, in this methodology, would have pointed either at inefficiency of the market proxy, or at the fact that betas are not very stable, or, related, at the fact that the *x* variables somehow help explain future betas or include information about some efficient portfolio (such as the "true" market portfolio).

6. PRICE ADJUSTMENT IN THE CAPM

(a) The Distribution of Asset Payoffs (Dividends) as a Basic Characterization

In the one-period formulation of the CAPM, an initial investment of one unit in asset *i* pays a gross return of $1 + r_i$. At a more basic level, however, one buys a (share in) project *i* at price P_i and this project pays a dividend at the end of the period, D_i. Thus,

(1) $1 + r_i = D_i / P_i \ , \quad 1 + \mu_i = E(D_i)/P_i \ .$

It is important to realize that the fundamental characteristics of the project (or set of projects in the case of a firm) are given by the distribution of D_i whereas the distribution of $1 + \mu_i$ depends also on the market environment that determines P_i. Thus it is more natural to formulate variance and covariance with the market associated with asset *i* in terms of D_i rather than $1 + \mu_i$. For easy of exposition in this section we will take the *distribution of the market return and the risk free rate* as given and hence the discussion here is of a partial equilibrium nature.

(b) Reformulation of the CAPM in Terms of Payoffs

We can write the basic CAPM equilibrium asset pricing equation as:

(2) $1 + \mu_i - (1 + r_f) = (\sigma_{im}/\sigma_m^2) \ [1 + \mu_m - (1 + r_f)].$

From equation (1) we can write the standard deviation, correlation and covariance with the market of the *payoff* of asset *i* as:

(3) $\sigma_i = \sigma_i^D / P_i \ , \quad \sigma_{im} = \mathrm{Cov}(D_i/P_i, r_m) = \sigma_{im}^D / P_i \ , \quad \rho_i \equiv \sigma_{im}/\sigma_i\sigma_m = \rho_i^D .$

Thus, the standard beta of asset *i*, the standard deviation of the return of asset *i*, the mean return of asset *i*, and the covariance of the return of asset *i* with the market return all depend on the price of asset *i*. On the other hand, the

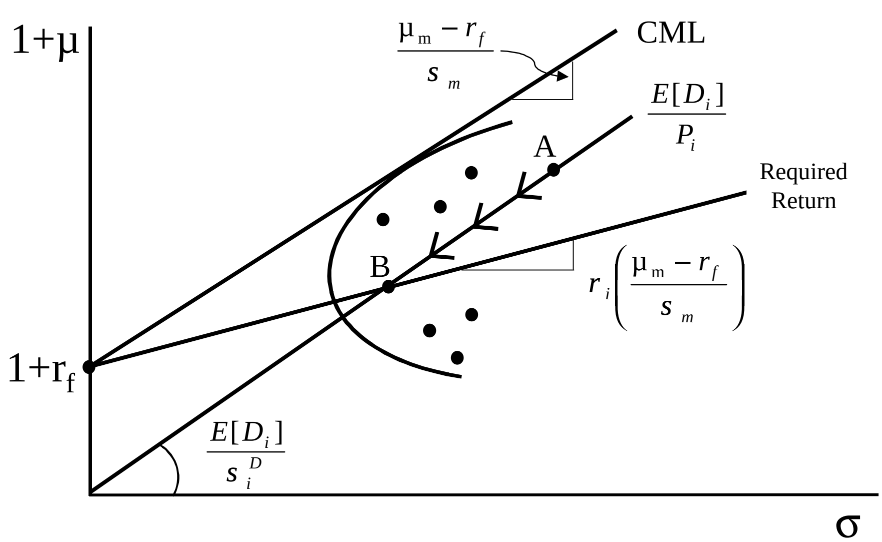

Figure 8
Price Adjustment in the CAPM
The price adjusts from A to B, where the expected return equals the required return.
Each asset in the opportunity set is equilibrium priced as a result of price adjustment.

correlation of the return of asset *i* with the market return and the ratio $(1+\mu_i)/\sigma_i$ do not depend on the price of asset *i*.

Multiply both sides of equation (2) by P_i:

$$(4) \qquad E(D_i) - P_i(1+r_f) = (\sigma_{im}^D/\sigma_m^2)\,(\mu_m - r_f).$$

The left-hand side of equation (4) indicates the expected payoff of asset *i* net of a risk-free opportunity cost. It is decreasing in the price of asset *i* so that demand for asset *i* depends negatively on its price. The right-hand side indicates the required payoff for asset *i* given as the product of the inherent systematic risk of asset *i* and the market risk premium. Equilibrium implies that the expected net payoff of asset *i* as a reward for taking the systematic risk associated with asset *i* is equal to the competitively determined "required" mean return for taking this systematic risk.

The price adjustment process inherent in equilibrium models is now shown explicitly for the CAPM. Suppose that initially the l.h.s. exceed the r.h.s. in equation (4) – the expected net payoff exceeds the required payoff. Then demand for asset *i* increases (perfectly elastically in a competitive market) and accordingly the price of asset *i* increases. This price continues to increase, increasing the opportunity cost of investment, until the expected net payoff is equalized with the required payoff and equation (4) holds.

To see the implications of price adjustment more sharply in terms of the opportunity set available to investors, rewrite equation (4) as follows:

$$(5) \qquad \frac{\mu_i - r_f}{\sigma_i} = \frac{1+\mu_i-(1+r_f)}{\sigma_i} = \frac{E(D_i)/P_i}{\sigma_i^D/P_i} - \frac{P_i(1+r_f)}{\sigma_i^D} = \rho_i\,\frac{\mu_m - r_f}{\sigma_m}\,.$$

The Sharpe Ratio for asset *i* must be equal to ρ_i times the Sharpe Ratio for the market. Adjustment of P_i ensures that the Sharpe Ratio for asset *i* adjusts to make this hold. The first term in the third expression indicates the ratio $(1+\mu_i)/\sigma_i$ which does not depend on P_i, even though $(1+\mu_i)$ and σ_i do individually.

Figure 8 illustrates the impact of price adjustment. With sigma on the horizontal axis and the expected gross return on the vertical axis, we have for asset *i*: $(1+\mu_i)/\sigma_i = E(D_i)/\sigma_i^D$ thus the ratio of mean and standard deviation of the return of asset *i* is constant as shown by the straight line emanating from the origin. Figure (3) also shows the CML with intercept equal to the gross risk free rate and with slope equal to the Sharpe ratio for the market and a line indicating the "required return" for asset *i* which has slope equal to ρ_i times the Sharpe ratio for the market. Suppose that the $(1+\mu_i, \sigma_i)$ "dot" is initially at point A (which must be on the line starting at the origin). Then the expected return exceeds the required return and the price of asset *i* rises. As a result the "dot" moves along the line towards the origin until it intersects with the required return line at point B. Note that price adjustment qualitatively does not guarantee market clearing. The reason is that the higher price lowers the mean return of the asset but also lowers the risk of the asset (less payoff risk per dollar invested). Quantitatively, however, equations (4) or (5) show that the net effect of the price increase on demand for the asset is negative.

(c) Applications and Exercises

1. Demonstrate the process of price adjustment as in Figure 8 but in $1 + \mu$, β space for a security whose payoffs are positively correlated with the market return. Do the same for a security whose payoffs are negatively correlated with the market return.

2. Show that no security with $(1 + \mu_i)/\rho_i\sigma_i < (\mu_m - r_f)/\sigma_m$ would ever be created and explain why.

3. Consider the following three risky assets. Asset 1 has $\mu_1 = 2$, $\sigma_{11} = 2$, and $\sigma_{12} = 1$, $\sigma_{13} = 0$. Asset 2 has $\mu_2 = 4$, $\sigma_{22} = 4$, and $\sigma_{21} = 1$, $\sigma_{23} = 0$. Asset 3 has $\mu_3 = 2$, $\sigma_{33} = 1$, and $\sigma_{31} = 0$, $\sigma_{32} = 0$.

   (a) Find the mathematical expression for the portfolio frontier. Illustrate graphically using the means and standard deviations of the individual assets as well as the portfolio frontier. You may, but need not, use a matrix approach.
   (b) Explain that the means, variances, and covariances stated above could not reflect a market equilibrium situation if the mean market return equals $\mu = 3½$.
   (c) Given a risk free asset with return $r_0 = 1$, obtain the tangency portfolio.

4. Explain for the equilibrium outcome in the zero-beta CAPM, whether it is possible for an individual investor to hold a negative quantity in an asset that is in positive aggregate supply.

* 7. GENERAL EQUILIBRIUM PRICE ADJUSTMENT IN THE CAPM

(a) Formal Derivation

Prices for all existing marketable assets, including the price of a risk free discount bond are determined endogenously. Consider the portfolio choices of individual investor *k* who determines the number of shares a_{ki} bought of each asset *i*. The initial wealth of the investor is spent as follows:

(1) $$\bar{w}_k = a_{kf}p_f + \boldsymbol{a_k^T p} ,$$

where the price and quantity of shares of the risk free discount bond is indicated by subscript *f* and vectors of risky asset prices and shares bought by investor *k* are indicated in bold face.

Expected end-of-period wealth is then given as:

(2) $$E(w_k) = a_{kf} + \boldsymbol{a_k^T} E(\boldsymbol{D}),$$

where the discount bond pays one unit of real wealth and each share of risky asset *i* pays D_i which is indicated again in bold face vector notation. Using equation (1) we can rewrite equation (2) as:

(3) $$E(w_k) = (\bar{w}_k/p_f) + \boldsymbol{a_k}^{\boldsymbol{T}}[E(\boldsymbol{D}) - (\boldsymbol{p}/p_f)].$$

Efficient portfolio choice implies choosing portfolio shares to minimize (half times) the variance of wealth, $Var(w_k)$:

(4) $$\underset{\boldsymbol{a_k}}{\text{Min}} \quad \tfrac{1}{2}\, \boldsymbol{a_k}^{\boldsymbol{T}} \Sigma\, \boldsymbol{a_k} \,,$$

subject to equation (3). Here Σ indicates the variance-covariance matrix of the *payoffs* of all assets. The first-order conditions are given as:

(5) $$\boldsymbol{a_k}^{\boldsymbol{T}} \boldsymbol{\Sigma} = \lambda_k [E(\boldsymbol{D}) - (\boldsymbol{p}/p_f)],$$

where λ_k represents the Lagrangian multiplier for the expected wealth constraint of investor *k*.

Now sum equation (5) over all investors. This yields:

(6) $$\boldsymbol{a}^{\boldsymbol{T}} \boldsymbol{\Sigma} = \lambda [E(\boldsymbol{D}) - (\boldsymbol{p}/p_f)] \,,$$

where $\lambda = \sum_k \lambda_k$ and $\boldsymbol{a}^{\boldsymbol{T}} = \sum_k \boldsymbol{a_k}^{\boldsymbol{T}}$. Market clearing for each asset implies that:

(7) $$\boldsymbol{a}^{\boldsymbol{T}} = \boldsymbol{1}^{\boldsymbol{T}}, \; a_f = 0.$$

This is true since all of the *shares* bought in a particular asset imply shares in the payoffs of the asset and must add up to one (which is the exogenously given supply of the asset; if the asset is a bigger firm, say, then the expected dividends will just be a larger amount). It is standard to assume that the risk free asset arises due to individuals providing loans and borrowing (without bankruptcy risk). Hence, the aggregate supply of the risk free asset must be zero. Thus, trivially, equation (6) becomes:

(8) $$\boldsymbol{1}^{\boldsymbol{T}} \boldsymbol{\Sigma} = \lambda [E(\boldsymbol{D}) - (\boldsymbol{p}/p_f)].$$

Postmultiply both sides of equation (8) by **1**. This yields:

(9) $$\boldsymbol{1}^{\boldsymbol{T}} \boldsymbol{\Sigma}\, \boldsymbol{1} = \lambda [E(w) - (\bar{w}/p_f)].$$

This follows since post-multiplying by **1** in the r.h.s. of equation (8) is equivalent to adding all payoffs and subtracting all initial prices deflated by the price of a discount bond. The expression in equation (9) then is found by adding equation (3) over all investors and evaluating at equilibrium using equation (7).

Since the l.h.s. of equation (9) is a scalar, we can eliminate the λ from equations (8) and (9) to yield:

(10) $$[E(\boldsymbol{D}) - (\boldsymbol{p}/p_f)]^T = \frac{\mathbf{1}^T \Sigma}{\mathbf{1}^T \Sigma\, \mathbf{1}} [E(w) - (\bar{w}/p_f)] \ ,$$

where $E(w) = E(\boldsymbol{D})\,\mathbf{1}$, as follows from equation (1) evaluated in equilibrium after aggregation. Taking the *i*th element from the vector yields:

(11) $$E(D_i) - (p_i/p_f) = \frac{\mathrm{Cov}(D_i, w)}{\mathrm{Var}\, w} [E(w) - (\bar{w}/p_f)]$$

Dividing both sides of equation (11) by p_i and dividing and multiplying the r.h.s. of equation (11) by $\bar{w}^2$ yields the standard CAPM equation if one considers that for a discount bond $1 + r_f = 1/p_f$.

(b) Discussion

Equation (10) consists of n independent asset pricing equations (instead of $n + 1$) like equation (11). The reason is that summing equation (10) over all i (by post-multiplying by **1**) does not produce an identity as we also implicitly incorporated equation (1) which leads to:

(12) $$\bar{w} = \mathbf{1}^T \boldsymbol{p} \ ,$$

after aggregation and in equilibrium. Thus, given the exogenous multivariate distribution of dividend payoffs, we can solve for all asset prices in terms of the *n+1*st asset, the risk free asset. The other equation, determining the risk free return, would come from an explicit derivation, after specifying preferences for all investors, which would essentially determine λ. Once we have λ, equation (9) would give the risk free rate. Of course λ may depend of many of the parameters of the model. However, for CARA preferences of all investors, it would basically be a constant being related to the inverse of some aggregate measure of absolute risk aversion. To see this last point in a more intuitive way consider that from the Lagrangian efficient portfolio problem of each investor we could derive that:

(13) $$1/\lambda_k = \frac{\partial E(w_k)}{\partial Var(w_k)} \ .$$

With this knowledge it is easier to interpret the equilibrium value of the risk free rate derived from equation (9):

(14) $$1 + r_f = 1/p_f = \frac{[E(\boldsymbol{D})^T \mathbf{1} \ - (\mathbf{1}^T \Sigma\, \mathbf{1}/\lambda)]}{\bar{w}} .$$

It is the "risk-adjusted return". Total dividends minus a compensation for risk per unit of invested wealth. The risk adjustment accounts for aggregate risk multiplied by a measure of aggregate risk aversion, $1/\lambda$.

When we fix the net number of discount bonds at zero, we use essentially an approach of Lucas (1978) where quantities of assets are taken exogenously. Alternatively, we may assume that a risk free commodity can be produced and fix the risk free rate exogenously. Then the number of risk free units produced will be determined endogenously. This approach is often associated with a paper of Cox, Ingersoll, and Ross (1981). Note by the way that, since the dividend distributions are fixed exogenously, we do not have a true general equilibrium model.