Chapter IV. Static Asset Pricing Models

While the CAPM has been, and despite its limitations, still is, the most popular asset pricing model, there is a variety of other asset pricing models that deserves attention. Many of these are variations of the basic CAPM. Here we will discuss only these models that are static in nature as is the basic CAPM.

1. THE CAPM WITH NON-MARKETABLE HUMAN CAPITAL

Here we drop assumption 6 of section III.1(b) that all assets owned by an investor are marketable. Clearly this assumption is violated for human capital. As "indentured servitude" is illegal nowadays in most countries, one is not allowed/able to sell one's future labor. Quantitatively, the impossibility of trading one's human capital may be quite important. Consider for instance a 30-year old, earning an annual salary of $50,000. Typically, this individual may have accumulated savings (investable non-human wealth) of no more than $100,000 including a pension plan, but excluding the equity in a home which cannot be invested. However, the real present value of the individual's life-time salary (human wealth) may be in the order of magnitude of $1,000,000. Thus, no more than around 10% of total wealth is marketable for this typical 30-year old.

Clearly, this 30-year old individual, in choosing how to invest the $100,000 of investable wealth, will not so much be concerned with market risk as with hedging the uncertainty in life-time earnings. On the other hand, a 65-year old will be much more concerned with market risk and may have very little remaining uncertainty in non-financial life-time earnings. It is then immediately clear that the Mutual Fund Theorem of the simple CAPM should fail. Not all individuals will hold the same portfolio of risky assets. We will here not be concerned with the exact implications for portfolio choice. Instead, we will ask the question of how dropping assumption 6 will affect asset pricing in equilibrium. Two questions present themselves. First, under what conditions will the individual differences due to disparities in non-tradable asset positions affect asset prices in the aggregate? The answer depends of course on the degree to which individual differences are systematic or idiosyncratic. Second, if asset prices are affected, does this present itself in the form of a different beta for the market factor, or will a second beta arise to deal with a non-market factor? In the following we will address these questions based on the work of Mayers (1972). See also Campbell (1996) and Jaganathan and Wang (1996).

For simplicity we use the method of proof of section III.2(b). Consider an individual k who maximizes the expected utility of end-of-period wealth given individual-specific non-marketable wealth (initial $\bar{w}_{nk}$, end-of-period w_{nk}) :

$$(1) \qquad \underset{\{s_{ik}\}_{i=0}^{n}}{\text{Max}} \; E[u_k(w_k)]$$

$$(2) \qquad \text{s.t.} \quad w_k = w_{nk} + \sum_{i=0}^{n} s_{ik}\,(1+r_i)\,\bar{w}_{mk}$$

$$(3) \qquad \text{s.t.} \quad \sum_{i=0}^{n} s_{ik} = 1 .$$

Here only equation (2) deviates from the original specification. Using the exact same derivation as in section III.2(b) (not reproduced here), we obtain:

$$(4) \qquad \mu_i - r_f = A \,\mathrm{Cov}\,(w, r_i) = A\,[\bar{w}_n \sigma_{in} + \bar{w}_m \sigma_{im}],$$

which follows since $w \equiv \Sigma_k w_k = w_n + w_m \equiv \bar{w}_n(1 + r_n) + \bar{w}_m(1 + r_m)$. We also define $A = [\sum_{k=1}^{K} \theta_k^{-1}]^{-1}$. Note that A in general depends on the distribution of wealth across individuals. When we use equation (4) for asset m we obtain:

$$(5) \qquad \mu_m - r_f = A\,[\bar{w}_n \sigma_{mn} + \bar{w}_m \sigma_m^2].$$

To find a fairly standard asset pricing equation we divide equations (4) and (5) to eliminate A. This yields:

$$(6) \qquad \mu_i - r_f = \beta_i^n (\mu_m - r_f), \quad \text{with} \quad \beta_i^n \equiv \frac{\sigma_{iw}}{\sigma_{mw}} \equiv \frac{\bar{w}_n \sigma_{in} + \bar{w}_m \sigma_{im}}{\bar{w}_n \sigma_{mn} + \bar{w}_m \sigma_m^2} .$$

Introducing non-marketable assets thus has several implications. As indicated previously, individuals will not hold the same portfolios of risky assets since they hold different types and quantities of non-marketable human capital. However, equation (6) indicates that nevertheless asset pricing is still independent of individual preferences. While idiosyncratic risk of the non-marketable asset will affect portfolio choice of the individual, it is only the systematic, economy-wide, component of non-marketable asset returns that matters. Asset pricing is still affected by covariance risk but it is now an asset's covariance with the market as well as its covariance with the systematic non-market asset return that matters.

A practical problem with equation (6) is that the beta from a standard regression now will no longer be equal to β_i^n. Instead, an *instrumental variable* regression, with true wealth w serving as the instrument for the market would provide exactly the right beta coefficient:

$$(7) \qquad \beta_i^n = \frac{\mathrm{Cov}(r_i, w)}{\mathrm{Cov}(r_m, w)} ,$$

as shown in Appendix B. An empirical test of this model can be found in Fama and Schwert (1977).

Rewriting equation (6) allows a simple intuitive interpretation:

$$(8) \qquad \mu_i - r_f = \lambda_n [\sigma_{im} + (\bar{w}_n/\bar{w}_m)\sigma_{in}], \quad \text{with} \quad \lambda_n \equiv \frac{\mu_m - r_f}{(\bar{w}_n/\bar{w}_m)\,\sigma_{mn} + \sigma_m^2} \equiv A\,\bar{w}_m .$$

Introduction of non-marketable assets causes two changes in the asset pricing equation. First, the price of market risk λ_n may not be affected as follows from equation (5), but it falls *for a given equity premium* when market and non-market returns are positively correlated (rises when negatively correlated). Second, an individual asset's risk now consists, aside from the usual co-movement with the market return, also of the covariance risk with the economy-wide non-market return, weighted by its relative importance. Thus, the systematic risk of non-market assets is priced even though these assets are not traded.

Why do we not obtain a two-beta formulation, or, similarly, a three-fund separation result? The issue is that the risk free asset and the market portfolio now are no longer sufficient to summarize the market opportunities of an individual investor. The investor also needs a fund to hedge against changes in the value of his non-market assets. For portfolio choice, the presence of non-marketable idiosyncratic risk means that no separation result can obtain. However, idiosyncratic risk at the market level is replaced by systematic risk of non-marketable assets which can be summarized by one additional factor.

Assume that an asset n exists with return perfectly correlated with the aggregate return on non-marketable assets. From equation (4) this will have excess expected return of:

$$\text{(9)} \qquad \mu_n - r_f = A\,[\bar{w}_n \sigma_n^2 + \bar{w}_m \sigma_{mn}].$$

Combining equations (5) and (9) to solve for $A\bar{w}_n$ and $A\bar{w}_m$ and substituting into equation (4), it is possible to write:

$$\text{(10)} \qquad \mu_i - r_f = \beta_{im}(\mu_m - r_f) + \beta_{in}(\mu_n - r_f)\,,$$

$$\text{where} \quad \beta_{im} = \frac{\sigma_{im}\sigma_n^2 - \sigma_{in}\sigma_{mn}}{\sigma_m^2\sigma_n^2 - \sigma_{mn}^2}\,, \qquad \beta_{in} = \frac{\sigma_{in}\sigma_m^2 - \sigma_{im}\sigma_{mn}}{\sigma_m^2\sigma_n^2 - \sigma_{mn}^2}\,.$$

The betas, β_{im} and β_{in}, may be obtained as the multiple regression coefficients in a regression of $r_i - r_f$ on a constant, $r_m - r_f$ and $r_n - r_f$. Jagannathan and Wang (1996), without derivation, use separate betas for the market return as well as for the return on human capital.

2. THE CAPM WITH MULTIPLE CONSUMPTION GOODS

In the basic static CAPM there is only one consumption good. As all end-of-period wealth is spent on this consumption good, the covariance with the marginal utility of consumption becomes covariance with wealth and the market return on wealth. In general, however, utility, even in a one-period model, will depend on the consumption of various consumption goods. We consider here the consequences of dropping assumption 3 in section III.1(b) of a composite consumption good.

For simplicity, consider two consumption goods: a composite good c which is the numeraire and represents regular consumption, and a good h which one may think of as housing and has relative price of p. Thus, for individual household k, we have the following decision problem:

(1) $$\max_{\{s_{ik}\}} \; E\left(\max_{c_k,\, h_k} \; [u_k(c_k, h_k)] \right)$$

(2) $$\text{s.t.} \quad c_k + p h_k = w_k \,,$$

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

The consumption allocation decision can be made are uncertainty is revealed; thus, the above problem can be decomposed into the following "two-stage budgeting" formulation:

(4) $$v_k(w_k, p) \equiv \max_{c_k,\, h_k} \; u_k(c_k, h_k)$$

(5) $$\text{s.t.} \quad c_k + p h_k = w_k \,.$$

Together with:

(6) $$\operatorname*{Max}_{\{s_{ik}\}} \; E[v_k(w_k, p)],$$

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

Here the decision problem in equations (4) and (5) serves only to determine the $v(\,)$ function. Clearly optimal c and h will be functions of the parameters exogenous to the household: w_k and p only. It is straightforward to check that $v_k(w_k, \hat{p}) = v_k(w_k/\hat{p})$, where $\hat{p} \equiv f(1, p)$, if $u_k(c_k, h_k)$ is *homothetic*. In this case the problem degenerates to the basic CAPM with real wealth as the only factor.

The first-order conditions for the optimization problem of equations (6) and (7) are:

(8) $$E[v_{k1}(w_k, p)\,(r_i - r_f)] = 0, \quad \text{for all } i \in \{1, n\},$$

where numerical subscripts j indicate partial derivatives with respect to the jth function argument. Using the definition of covariance [see Appendix] we obtain:

(9) $$E[v_{k1}(w_k, p)](\mu_i - r_f) = -\text{Cov}[v_{k1}(w_k, p)\,, r_i].$$

If we assume that returns are normally distributed so that w_k is normal and that p is normally distributed, we can apply my modest *generalization of Stein's Lemma* [see Appendix C] stating that, when x, y, and z are multivariate normal, then:

$$\text{Cov}[x, h(y, z)] = E[h_1(y, z)]\text{Cov}(x, y) + E[h_2(y, z)]\text{Cov}(x, z) \,.$$

Thus, using the lemma on equation (9):

(10) $$\mu_i - r_f = \frac{-E[v_{k11}(w_k,p)]}{E[v_{k1}(w_k,p)]}\,\mathrm{Cov}(w_k, r_i) + \frac{-E[v_{k12}(w_k,p)]}{E[v_{k1}(w_k,p)]}\,\mathrm{Cov}(p, r_i).$$

Equation (10) suggests a two-factor CAPM result. However, the expression includes various terms that are specific to individual *k*. The next step thus is to consider market equilibrium by aggregating over all individuals:

(11) $$\sum_{k=1}^{K} \frac{-E[v_{k1}(w_k,p)]}{E[v_{k11}(w_k,p)]}(\mu_i - r_f) =$$

$$\mathrm{Cov}(w_m, r_i) + \sum_{k=1}^{K} \frac{E[v_{k12}(w_k,p)]}{E[v_{k11}(w_k,p)]}\,\mathrm{Cov}(p, r_i),$$

which follows since $w_m = \Sigma_k w_k$. Next, consider that $w_m = \bar{w}_m(1 + r_m)$ and assume that an asset exists with return r_p that is *perfectly correlated* with *p* so that $p = \gamma + \delta r_p$. Then:

(12) $$\mu_i - r_f = A_1 \sigma_{im} + A_2 \sigma_{ip},$$

Where as before $\mathrm{Cov}(r_i, r_j) \equiv \sigma_{ij}$ and where:

$$A_1 = \bar{w}_m \Big/ \left[\sum_{k=1}^{K} \frac{-E[v_{k1}(w_k,p)]}{E[v_{k11}(w_k,p)]}\right] > 0 ,$$

$$A_2 = \delta\left[\sum_{k=1}^{K} \frac{E[v_{k12}(w_k,p)]}{E[v_{k11}(w_k,p)]}\right] \Big/ \left[\sum_{k=1}^{K} \frac{-E[v_{k1}(w_k,p)]}{E[v_{k11}(w_k,p)]}\right].$$

Note that δ appears in A_2 because we converted from covariance between return on asset *i* and *p* to covariance between return on asset *i* and the return on the asset perfectly correlated with *p*. The sign of A_2 depends negatively on the sign of v_{k12}.

Applying equation (12) to asset *m* (it is easy to check that, if equation (8) holds for any "primitive" asset *i* , it also hold for any portfolio, including the market portfolio):

(13) $$\mu_m - r_f = A_1 \sigma_m^2 + A_2 \sigma_{mp}.$$

Similarly, for the asset with return perfectly correlated with *p*,

(14) $$\mu_p - r_f = A_1 \sigma_{pm} + A_2 \sigma_p^2 .$$

Use equations (13) and (14) to solve for A_1 and A_2 :

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} \sigma_m^2 & \sigma_{mp} \\ \sigma_{mp} & \sigma_p^2 \end{pmatrix}^{-1} \begin{pmatrix} \mu_m - r_f \\ \mu_p - r_f \end{pmatrix} .$$

This yields:

(15) $$A_1 = \frac{\sigma_p^2(\mu_m - r_f) - \sigma_{mp}(\mu_p - r_f)}{\sigma_p^2 \sigma_m^2 - \sigma_{mp}^2} ,$$

$$A_2 = \frac{\sigma_m^2(\mu_p - r_f) - \sigma_{mp}(\mu_m - r_f)}{\sigma_p^2 \sigma_m^2 - \sigma_{mp}^2} .$$

Substitute equation (15) into equation (12) to obtain the expected return of any asset i as:

(16) $$\mu_i - r_f = \beta_{im}(\mu_m - r_f) + \beta_{ip}(\mu_p - r_f) ,$$

with: $$\beta_{im} = \frac{\sigma_p^2 \sigma_{im} - \sigma_{mp}\sigma_{ip}}{\sigma_p^2 \sigma_m^2 - \sigma_{mp}^2} , \quad \beta_{ip} = \frac{\sigma_m^2 \sigma_{ip} - \sigma_{mp}\sigma_{im}}{\sigma_p^2 \sigma_m^2 - \sigma_{mp}^2} .$$

Notice that the betas are simply the slope coefficients that would arise in a multi-variate regression of r_i - r_f on r_m - r_f and r_p - r_f .

The intuition of equation (16) is that investors care about the risk to their wealth. However, they separately also care about what they can do with their wealth. Thus, they care about *hedging* against changes in p (which you may think of as the price of housing, for instance). Typically, an increase in p would decrease the marginal (indirect) utility of wealth; i.e., $v_{k12} < 0$. Thus $A_2 > 0$. Equation (14) then implies that $\mu_p - r_f > 0$ as long as $\sigma_{mp} \geq 0$: the asset perfectly correlated with p is risky, since a high return is associated with a low marginal utility of wealth; it thus offers a positive risk premium in equilibrium.

The case considered here, where considering various types of consumption goods implies additional beta factors in the asset pricing equation, even in a static model, appears to have been largely overlooked in the literature. The only reference I have found in the literature where a bundle of consumption goods is considered and is shown to imply additional beta factors is a section in Breeden (1979, section 7) in the context of a continuous-time dynamic model. Lyon (2000) provides an interesting application of the material considered here where p represents the real price of housing.

Applications and exercises

1. Derive equation (15) for the case where no asset exists that is perfectly correlated with *p* [provide hint].

2. Explain the difference in implications of considering housing as a non-marketable asset discussed in section 2, or of considering housing as a key consumption good separate from other consumption goods, as discussed in section 3.

3. Discuss the signs of A_1 and A_2 by using information from the optimization problem in equations (4) and (5).

3. THE INTERNATIONAL CAPM

In which manner can we extend the CAPM to consider more than one country? One obvious modification is that covariance of a return with the market now becomes covariance of an asset's return with the return on a broader portfolio such as the world portfolio. As we will see, however, the standard CAPM requires in general a more substantial modification. Below, we discuss what is essentially a discrete-time version of the Adler and Dumas (1983) International CAPM (sometimes confusingly called the ICAPM which is the name typically reserved for the Intertemporal CAPM). For more recent work on international asset pricing in the same vein, see Black (1990), Stulz (1994), and DeSantis and Gérard (1998).

(a) Model Set-Up

We model a multi-country world. Capital flows are unrestricted. This implies that the law of one price prevails for all financial assets. That is, nominal returns on an asset *i* are equal for the residents in all countries. Trade flows are, however, costly and this implies that the law of one price does not prevail for the one consumption good we consider.[1] Thus, while asset returns are equalized across countries, purchasing power parity need not prevail.

To simplify matters further we assume just one representative consumer in each country. The consumer in benchmark country *J* faces the following standard decision problem:

$$(1) \qquad \max_{\{s_{iJ}\}_{i=0}^{n}} E[u_J(w_J)]$$

[1] Alternatively, we may think of there being a traded good and a non-traded good. The law of one price holds for the traded good but not for the non-traded good of each country. And so the price index for aggregate consumption as made up of the two goods may differ between the two countries. Note that a price index for the two goods may exist only if preferences for the two goods are homothetic which we assume to avoid the complications discussed in the previous section.

$$(2) \qquad \text{s.t.} \quad w_J = \sum_{i=0}^{n} s_{iJ}\,(1 + r_i)\,\bar{w}_J$$

$$(3) \qquad \text{s.t.} \quad \sum_{i=0}^{n} s_{iJ} = 1 .$$

Here the first subscript *i* indicates assets and the second subscript indicates the country of the consumer; all other notation is as before. Asset returns are in real terms:

$$(4) \qquad 1 + r_i = \frac{1 + r_i^{nJ}}{1 + \pi_J},$$

and measured in the currency of the benchmark country *J*. Thus r_i^{nJ} represents the nominal asset return in the currency of the benchmark country and π_J stands for the inflation rate in the benchmark country; both are unknown at the time of the portfolio choice.

The representative consumer in any other country faces the same opportunity set as does the benchmark consumer in the sense that the available assets and their returns are the same for both. This is the assumption of the law of one price as applied to all financial assets. However, the law must apply to *nominal* prices and returns only. Real asset prices and returns differ as the consumers in different countries face different consumer price indices and inflation rates.

The decision problem for any consumer outside of the benchmark country (that is countries *j*, where $j \in \{1, ..., J-1\}$, becomes:

$$(5) \qquad \underset{\{s_{ij}\}_{i=0}^{n}}{\text{Max}} \; E[u_j(w_j\, x_j)]$$

$$(6) \qquad \text{s.t.} \quad w_j = \sum_{i=0}^{n} s_{ij}\,(1 + r_i)\,\bar{w}_j$$

$$(7) \qquad \text{s.t.} \quad \sum_{i=0}^{n} s_{ij} = 1 ,$$

and where

$$(8) \qquad x_j \equiv \frac{1 + \pi_J}{1 + \pi_j}\, e_j ,$$

with e_j representing the rate of appreciation of country *j*'s currency in terms of the benchmark country's currency, $e_j \equiv E_j / \bar{E}_j$, with E_j is measured in units of country *j* currency per unit of country *J* currency. Equation (8) thus gives the real rate of appreciation of country *j*'s currency in terms of the benchmark country's currency. Or, equivalently, the rate of appreciation of country *j*'s real exchange rate. Real wealth w_j is measured in terms of the benchmark's country consumption basket as follows from equation (6) (given that returns r_i are defined in real benchmark country currency). Accordingly, to convert to country *j* purchasing power, we need to multiply by $1+\pi_J$ to get to nominal benchmark country terms, then multiply by e_j to convert to country *j* currency terms and lastly divide by $1+\pi_j$ to obtain real wealth in terms of purchasing power in country *j*. As follows from equations (5) and (8) these conversions are equivalent to multiplying w_j by the rate of appreciation of the real exchange rate x_j. Clearly, equations (5) - (8) apply to the benchmark country as long as one realizes that $x_J \equiv 1$.

We will take the x_j to be normally distributed as are the real returns r_i ; thus, the w_j are normally distributed as well. Of course, these assumptions imply that the real returns for the non-benchmark country investor are not normally distributed, but this will not present any technical complications.

Next we obtain the asset pricing equations for all assets phrased in terms of their purchasing power in the benchmark country. For empirical purposes, then, all returns are measured in real terms of the benchmark currency.

(b) Model Solution

Assume that an asset exists denominated in the benchmark currency that is riskless in real terms. Then the first-order conditions for the representative consumers become:

(9) $$E[u_{j1}(w_j x_j)\, x_j\, (r_i - r_0)] = 0, \quad \text{for all } i \in \{1, ..., n\},\ j \in \{1, ..., J\}\ .$$

Using first the definition of covariance and then applying Stein's Generalized Lemma yields for all *i* and *j*:

(10) $$(\mu_i - r_0)\, E[u_{j1}(w_j x_j)\, x_j] = -E[u_{j11}(w_j x_j)\, x_j^2]\, Cov(w_j, r_i)$$
$$- E[u_{j11}(w_j x_j)\, w_j\, x_j + u_{j1}(w_j x_j)]\, Cov(x_j, r_i)\ .$$

Rewriting the above expressions gives:

(11) $$(\mu_i - r_0)\, A_j = Cov(w_j, r_i) + B_j\, Cov(x_j, r_i)\ ,$$

where the constants are defined as:

(12) $$A_j = \frac{-E[u_{j1}(w_j x_j)\, x_j]}{E[u_{j11}(w_j x_j)\, x_j^2]}\ , \quad B_j = \frac{-E[u_{j11}(w_j x_j)\, w_j\, x_j + u_{j1}(w_j x_j)]}{E[u_{j11}(w_j x_j)\, x_j^2]}\ .$$

Define world real wealth in terms of the home consumption basket as *w* and the return on the world market portfolio as $1 + r_w = w/\bar{w}$. Then add equation (11) for all *j* to aggregate over the representative consumers and use the world market return definition to obtain:

$$(13) \qquad \mu_i - r_0 = \frac{\bar{w}\ Cov(r_w, r_i)}{\sum_{j=1}^{J} A_j} + \frac{\sum_{j=1}^{J-1} B_j\ Cov(x_j, r_i)}{\sum_{j=1}^{J} A_j} .$$

Notice that $Cov(x_J, r_i) = 0$ since $x_J = 1$.

Employing the by now familiar procedure:

$$(14) \qquad \mu_w - r_0 = \frac{\bar{w}\ Var(r_w)}{\sum_{j=1}^{J} A_j} + \frac{\sum_{j=1}^{J-1} B_j\ Cov(x_j, r_w)}{\sum_{j=1}^{J} A_j} ,$$

$$(15) \qquad \mu_{x_k} - r_0 = \frac{\bar{w}\ Cov(r_w, x_k)}{\sum_{j=1}^{J} A_j} + \frac{\sum_{j=1}^{J-1} B_j\ Cov(x_j, x_k)}{\sum_{j=1}^{J} A_j} ,$$

for all $k \in \{1, .., J-1\}$.

Note that $1 + \mu_{x_j} = x_j$.

Using our standard notation $\sigma_{ij} = Cov(r_i, r_j)$, solution of (14) and (15) and substitution into (13) analogously to that in section 3 can be applied to find the International CAPM equation. Specifically, the derivation is as follows. Set

$$C_j = \frac{B_j}{\sum_{j=1}^{J} A_j}, \quad j = 1, ..., J-1, \qquad C_w = \frac{\bar{w}}{\sum_{j=1}^{J} A_j},$$

then we can combine equations (14) and (15) in matrix notation as:

$$(16) \qquad \boldsymbol{\mu_x} - \boldsymbol{r} = \boldsymbol{\Sigma_x C},$$

where:

$$\boldsymbol{\Sigma_x} = \begin{pmatrix} \sigma_{x_1 x_1} & \sigma_{x_1 x_2} & \cdots & \sigma_{x_1 w} \\ \sigma_{x_2 x_1} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{w x_1} & \cdots & \cdots & \sigma_{ww} \end{pmatrix}, \quad \boldsymbol{\mu_x} - \boldsymbol{r} = \begin{pmatrix} \mu_{x_1} - r_0 \\ \vdots \\ \mu_{x_{J-1}} - r_0 \\ \mu_w - r_0 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_w \end{pmatrix}.$$

A matrix version of equation (13) is:

$$\text{(17)} \qquad \boldsymbol{\mu_i} - \boldsymbol{r} = \boldsymbol{\Sigma_i C},$$

where

$$\boldsymbol{\Sigma_i} = \begin{pmatrix} \sigma_{1x_1} & \sigma_{1x_2} & \cdots & \sigma_{1w} \\ \sigma_{2x_1} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{nx_1} & \cdots & \cdots & \sigma_{nw} \end{pmatrix}, \quad \boldsymbol{\mu_i} - \boldsymbol{r} = \begin{pmatrix} \mu_1 - r_0 \\ \vdots \\ \vdots \\ \mu_n - r_0 \end{pmatrix}$$

Solving for $\boldsymbol{C}$ from equation (16) and substituting into equation (17) produces:

$$\text{(18)} \qquad \boldsymbol{\mu_i} - \boldsymbol{r} = \boldsymbol{\Sigma_i} \boldsymbol{\Sigma_x^{-1}} (\boldsymbol{\mu_x} - \boldsymbol{r}).$$

The expression $\boldsymbol{\Sigma_i} \boldsymbol{\Sigma_x^{-1}}$ is recognizable as the (transpose of) the expression for the theoretical slope coefficients in a standard OLS regression (without a constant).[2] Thus, we can write for any asset *i*:

$$\text{(19)} \qquad \mu_i - r_0 = \sum_{j=1}^{J-1} \beta_{i x_j} (\mu_{x_j} - r_0) + \beta_{iw} (\mu_w - r_0),$$

(c) Interpretation

The extension of the CAPM to a multi-country case leads to a *J*-factor solution for the pricing of assets. The new factors are the excess returns on the *J-1* assets that are perfectly correlated with the real exchange rate appreciations x_j for each country but the benchmark country. Under global purchasing power parity, x_j would be deterministic for all *j* and equal to one so that we would be back to a one-factor model. More generally, and in particular when nontraded

[2] In particular, we can identify $\boldsymbol{\Sigma_x}$ with $\boldsymbol{X'X}$ and $\boldsymbol{\Sigma_i^T}$ with $\boldsymbol{X'Y}$.

goods exist, the x_j will vary and may have an expected value that is different from one. One can get real return x_j by holding an uncovered position in the currency of country *j* ; so there is a clear empirical measure and interpretation of the factor x_j.

Why are there these extra factors affecting expected returns? From a foreign (i.e. not benchmark country) consumer's perspective, utility varies not just because of variation in wealth but also because of variation in the purchasing power of the wealth. The foreign consumer may *hedge* against changes in the "terms of trade" by holding her own currency. If the foreign currency appreciates, then, for given market returns (denominated in the domestic currency), the purchasing power of returns has diminished but this is offset in part by the gain in having held the foreign currency. For the domestic (benchmark) consumer there is no need to hold the foreign currency. However, it is priced as an independent factor due to the foreign consumer. Typically, the domestic consumer would not hold the foreign currency or might short the foreign currency. Clearly, the two-fund separation result implying that all consumers hold risky assets in the same proportions does not hold in this case.

To see why the standard CAPM breaks down consider the Efficient Frontier and Capital Market Line for real returns (measured in the investor's own currency) as displayed in Figure 2. If we are to consider indifference curves in the space of mean returns and standard deviation of returns then the real returns must be measured in terms of each consumer's relevant price level. But this implies that the opportunity sets for each consumer are different: for equal nominal returns, the real returns will vary by consumer. Thus in any two countries *A* and *B* we would have two different Capital Market Lines (with, possibly, both different slope and intercept) and so the simple CAPM would not apply. Interestingly, the fact that apparently identical opportunity sets for different consumers imply different opportunity sets in practice, similarly applies to the case of heterogeneous expectations! Thus, in mathematical terms, a model with heterogeneous expectations would be identical to our present model with heterogeneous consumption opportunities.

Two questions need to be answered concerning the nature of a real exchange rate factor: the sign of its beta and the sign and level of its premium. To simplify the discussion, assume that the covariance between the real growth

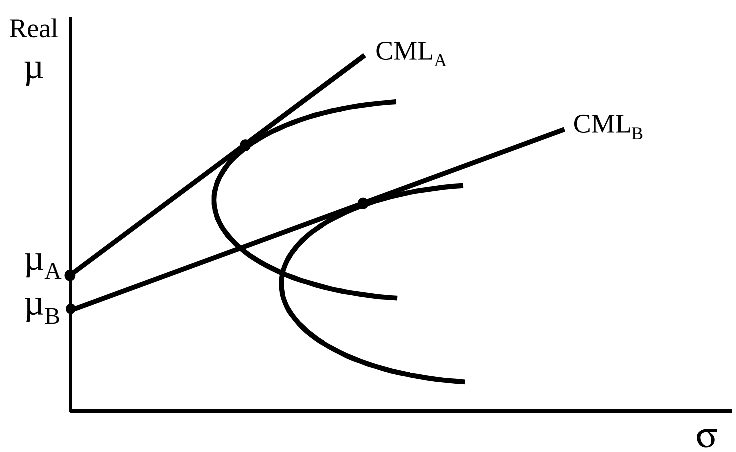

Figure 1
Opportunity Sets with Exchange Rate Risk
Real returns may vary by consumer, possibly resulting in two different Capital Market Lines and opportunity sets for each consumer A and B.

rate of global wealth (the world market return) and the real return on each foreign currency is zero and that the real currency returns are also mutually uncorrelated. Then each slope in equation (19) becomes a simple regression slope whose sign is determined by its covariance with asset *i*'s return only. The covariance between the foreign currency return and asset *i* could be of either sign. For instance, if the asset represents a nontraded good produced in the foreign country then its return will be high if there is high productivity in its production. But this may also imply a lower price for the nontraded good and a lower price level in the foreign country, raising the real exchange rate *x*. Thus, in this example, the beta for this asset would be positive.

More interesting is the question about the sign of the risk premium on factor x_j. Assume again that the covariances between the foreign currency return and all other foreign currency returns, as well as the global market return are zero. Then equation (15) implies that the sign of $\mu_{x_j} - r_0$ is given by B_j, which by equation (12) is given by $\partial E[u_{j1}(w_j x_j)\, x_j]/\partial x_j = E[u_{j11}(w_j x_j)\, w_j x_j + u_{j1}(w_j x_j)]$. This is ambiguous in sign showing two opposing effects: an increase in x_j directly raises the real return (by raising the purchasing power to the foreign consumer) thus raising the marginal valuation of wealth; a change in x_j also raises real wealth thus lowering the marginal utility of wealth. Equation (15) also shows that the risk premium on factor x_j is larger as the variance of the return is larger; further, the premium is smaller the more agents exist outside the foreign country, as reflected in the sum of the A_j.

The model of Adler and Dumas (1983) differs from the current model in some respects. First it is solved in continuous time. Second, it is formulated in nominal terms which appears to be incorrect when inflation rates are stochastic. Otherwise, our result is as in Adler and Dumas that *n-1* factors exist outside of the world market return related to exchange rate risk for all currencies/countries relative to a benchmark.

4. ARBITRAGE PRICING THEORY

Arbitrage Pricing Theory, or the APT, is a competitor to the CAPM and explains asset prices in a way that is fundamentally different. Developed by Ross (1976), it employs a type of arbitrage that differs from standard arbitrage in that it relies on diversification to reduce all (or almost all) risk. It differs from the CAPM in set-up by not relying on elliptical distributions or quadratic utility but, instead, assumes a linear error structure.

(a) Model Set-Up

The (net) expected return on any asset *i* is given tautologically as:

$$(1) \qquad r_i \equiv \mu_i + \eta_i \,, \qquad E(\eta_i) \equiv 0 \,.$$

The key restriction is on the error term which is assumed to have the following "factor" structure:

$$(2) \qquad \eta_i = b_{i1}\tilde{F}_1 + b_{i2}\tilde{F}_2 + ...b_{ik}\tilde{F}_k + \varepsilon_i \,,$$

$$E(\tilde{F}_i) = E(\varepsilon_i) = E(\varepsilon_i \tilde{F}_j) = E(\varepsilon_i \varepsilon_h) = 0.$$

The K "factors" $\tilde{F}_j$ have a mean of zero. The strength of the effect of each factor j on r_i (or equivalently on η_i) is given by the "factor loadings" b_{ij} . The errors ε_i are uncorrelated with the factors and uncorrelated across assets.

Apart from the assumptions on the error structure, the APT assumes perfect markets (perfect competition and no frictions) and homogeneous expectations as does the CAPM. It also assumes that the number n of assets considered is much larger than the number K of factors. Importantly, since construction or identification of a market portfolio is not required in the APT, there is no reason to study the universe of assets. So even though $n >> K$ is required n need only be a subset of the assets in existence. For instance, the APT should hold for all Belgian assets of companies older than ten years, just as well as it would hold for all assets on the NYSE.

Define the concept of an *Arbitrage Portfolio* as a portfolio with (a) no wealth invested, i.e.,

$$(3) \qquad \sum_{i=1}^{n} s_i = 0;$$

and (b) no risk. The return on such an arbitrage portfolio r_p is given as:

$$(4) \qquad r_p = \sum_{i=1}^{n} s_i r_i = \sum_{i=1}^{n} s_i \mu_i + \sum_{k=1}^{K} (\sum_{i=1}^{n} s_i b_{ik}) \tilde{F}_k + \sum_{i=1}^{n} s_i \varepsilon_i \ .$$

Let's indicate the second term on the right-hand side of equation (4) by *systematic* risk and the last term by *nonsystematic* risk for obvious reasons. Now consider how to build the arbitrage portfolio to eliminate both types of risk:

$$(5a) \qquad \sum_{i=1}^{n} s_i b_{ik} = 0 \ , \qquad \text{for all } k.$$

$$(5b) \qquad n = \text{large} \ , \qquad |s_i| \approx 1/n \ .$$

Equation (5a) states the condition for eliminating all systematic risk; equation (5b) states the condition for eliminating all nonsystematic risk, relying on the law of large numbers.[3] Clearly, not all nonsystematic risk can be eliminated without having a portfolio existing of infinitely many assets. Thus, the "arbitrage" considered here is not fully riskless, but close enough so for practical purposes. Much of the literature on the APT deals with the last issue: how to eliminate the nonsystematic risk. Since the basic idea of this is clear we will not dwell on this issue.

From equations (4) and (5) we can now infer that:

$$(6) \qquad r_p = \sum_{i=1}^{n} s_i \mu_i = 0 \ .$$

The second equality follows from the "no arbitrage" condition: in perfect markets, no arbitrage opportunities should be available, thus the return on an arbitrage portfolio must be zero.

[3] It is not necessary the s_i is close to *1/n* (in absolute value). For instance, if we set $s_i = i/x$, where x is picked to make the shares add to one, then we can let the nonsystematic risk go to zero as $i \to \infty$.

(b) Model Solution

State equations (3), (5a), and (6) in matrix notation:

$$(7) \qquad \boldsymbol{s}^T (\boldsymbol{B} - \bar{\boldsymbol{R}}) = \boldsymbol{0} \,,$$

where: $\boldsymbol{s}$ is a *1xn* column vector of arbitrage portfolio shares; $\bar{\boldsymbol{R}}$ is an *n x (K+1)* matrix of *K+1* identical column vectors of the *n* expected returns on the assets in the arbitrage portfolio; and $\boldsymbol{B}$ is an *n x (K+1)* matrix which has *ones* in its first column and the factor loadings b_{ik} for the remaining K columns; **0** is a *n x 1* column vector consisting of zeros.

From basic linear algebra we know that the matrix $\boldsymbol{B} - \bar{\boldsymbol{R}}$ does not have full column rank. As a result we can write $\bar{\boldsymbol{R}}$ as a linear combination of the columns in $\boldsymbol{B}$:

$$\gamma_0 (1 - \mu_i) + \gamma_1 (b_{i1} - \mu_i) + \gamma_2 (b_{i2} - \mu_i) + ... \gamma_K (b_{iK} - \mu_i) = 0 \,,$$

which holds for all i and for some value of the γ_k . Rewriting the above equation produces:

$$(8) \qquad \mu_i = \lambda_0 + \sum_{k=1}^{K} \lambda_k b_{ik} \quad , \qquad \lambda_i = \gamma_i / \sum_{j=0}^{k} \gamma_j \ \text{ for all } i \,.$$

Note that the sum of the $\lambda_i = 1$. Equation (8) is the asset pricing equation of the APT.

(c) Discussion and Intuition

The expected return on an asset $\mu_i \equiv E(r_i)$ thus equals a linear combination of the loadings on the *K* factors. It is possible, however, to provide a more specific interpretation. If a riskless asset exists with return r_0 then it must be that the b_{0k} are all zero and so equation (8) implies:

$$(9) \qquad r_0 = \mu_0 = \lambda_0 \,.$$

Consider an asset *k* that has unit sensitivity to factor *k* and zero sensitivity to all other factors (such an asset can always be created if $\boldsymbol{B}$ is invertible). Then the expected return on asset *k* from equation (8) and using equation (9) equals:

$$(10) \qquad \mu_k = r_0 + \lambda_k \quad \rightarrow \quad \lambda_k = \mu_k - r_0 \,.$$

Thus, equation (8) becomes:

$$(11) \qquad \mu_i - r_0 = \sum_{k=1}^{K} b_{ik} (\mu_k - r_0) \,,$$

which is a multi-beta version of the CAPM pricing equation. Given equations (1) and (2), equation (11) can be written in *ex post* form:

$$r_i - r_0 = \sum_{k=1}^{K} b_{ik} (r_k - r_0) + \varepsilon_i . \tag{12}$$

Thus, if one knew the *K* factors, the factors loadings (or "betas") could be obtained econometrically as the multiple regression coefficients, with historical excess returns on the *K* factors as the independent variables and the historical excess return on asset *i* as the dependent variable.[4]

Figure 2 presents a graph from Roll and Ross (1980) which illustrates in a simple 1-factor example how absence of arbitrage implies linear pricing as in equation (11). The graph depicts the "security market line" for the 1-factor model. Suppose assets 1 and 3 are on the line and that asset 2 is above the line, meaning that it has a positive "alpha". Then one could construct a portfolio of assets 1 and 3 with identical systematic risk as asset 2 but with a lower expected return. By short selling the portfolio of assets 1 and 3 and buying asset 2 an arbitrage portfolio can be created (no investment but positive return, and no risk if we may ignore idiosyncratic risk). Arbitrage opportunities of this sort will be absent only if all assets lie along the line.

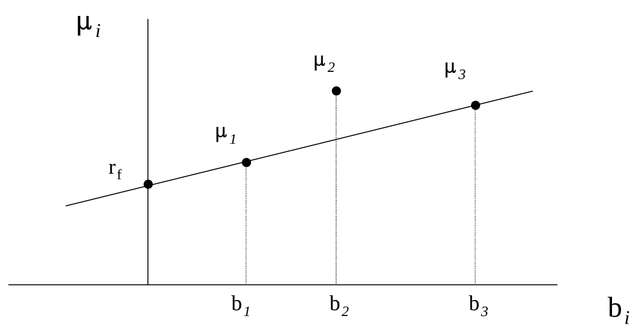

Figure 2
Arbitrage Pricing
Constructing a portfolio short in assets 1 and 3 and long in asset 2 can create an arbitrage portfolio without (systematic risk) but with positive payoff. Absence of arbitrage opportunities of this type guarantees that all assets lie on the line.

(d) Empirical Issues

Of course, the factors are not known and the APT does not provide any specific economic guidance on how to pick factors. In practice, purely statistical procedures are typically used to find the factors. Roll and Ross (1980) and

[4] Note that, due to idiosyncratic risk, it is not necessary that the APT will hold in the case of each individual asset; Specifically, there is no requirement that an idiosyncratic component with positive expected value is arbitraged away.

others use “factor analysis” or “principal components” (statistical techniques often used outside of finance and economics) to identify portfolios of assets that would have best explained the realized asset returns. The question is then: How many portfolios have a significant explanatory power for asset returns? Having obtained all significant portfolios (the “factors”) one then checks if any other, interesting variables have additional explanatory power. This becomes the formal test of the APT. In particular, say that there are three significant factors, then the question is, does the CAPM beta or total variance of return contribute any explanatory power in addition to the three statistical factors. The APT says no. In reality it seems that firm size has some additional explanatory power. This depends on the sample however: the APT allows any large set of assets to be investigated and the extent to which returns in a particular set of assets deviate from APT predictions may vary. As in the CAPM, a possible test of the APT is to consider whether the “alphas” are significantly different from zero.

In factor analysis or principal components analysis factors are picked in such a way that they are orthogonal. This is always possible. Say that two factors matter, X and Y. Then if these factors are correlated we can always decompose $Y = a + bX + g$, where $E(Xg) = E(g) = 0$, and then redefine X and g as the two, now orthogonal, factors. Clearly, in the case of orthogonal factors, the regression slopes in equation (12) would just be simple regression coefficients.

Chen, Roll, and Ross (1983) consider a large group of macroeconomic variables that could potentially be factors in pricing financial assets. They find four macro variables that are significant: industrial production; changes in the default risk premium on corporate bonds; changes in the term premium on long-term versus short-term bonds; and unanticipated inflation. The first variable may relate to profitability while the other three deal with the opportunity cost of holding stock (or the discount rate).

In the context of a macroeconomic model, the different random shocks that drive the equilibrium state of the model each should serve as a separate factor. Thus, in a real business cycle mode for instance, technology shocks and productivity shocks should be the factors that should price each financial asset. It appears that little empirical work has been done along these lines. This is unfortunate because, in the absence of a specific general equilibrium model, the APT is pretty much like an empty shell: it states that any “factor” could price financial assets but it does not limit in any way what the factors should be.

(e) Empirical Procedures

In cookbook style summary, here are the steps needed to test the APT via factor analysis:

1. Take any large group of financial assets over any returns horizon (daily is fine).
2. Find the Variance-Covariance matrix for realized returns from historical data.
3. Statistically use factor analysis to find orthogonal factors, using maximum likelihood analysis to determine the cutoff for the last significant factor.
4. Find the factor loadings b_{ik} in the time series.
5. Find the factor risk premia 8_k cross-sectionally.
6. See if all assets lie on the Arbitrage Pricing Line and/or see if other non-APT risk factors (such as the market beta) are priced.

The results are typically around four significant factors (without economic interpretation). Other variables (such as total risk) are irrelevant although this is debatable for firm size. The APT explains a significant part of the CAPM residuals.

As a specific example, the Roll and Ross (1980) paper considers daily data from July 1962 on. They construct alphabetically 42 groups of 30 assets, estimate for each group the variance-covariance matrix (the groups are formed for computational issues, to keep the covariance matrix manageable), and then apply a two-pass method. Since data are daily, estimation can be quite precise so it is not necessary to worry about measurement error as in the Fama-MacBeth approach. They then obtain simultaneously by factor analysis the factors and the loadings. Subsequently, factor risk premia are found cross-sectionally based on mean returns over the sample period (this is done through GLS to correct for the fact that the distribution of the error estimates depends on the factor loadings). Lastly, the size and significance of the factor risk premia is obtained. This is done separately for all 42 portfolios.

Chen (1983) employs a slightly different approach. He uses odd days in the sample to estimate the factor loadings in the first pass; then uses even days to test the APT in the second pass.

(f) Comparison with the CAPM

As stated previously, the APT does not require an assumption on the returns distribution; it does not require that the whole universe of assets be considered; and it has no specific role for the market portfolio. Its key assumption is the assumption of error terms that are linear in the relevant factor shocks. The framework is extremely flexible and in fact is not limited to a static interpretation: asset sensitivities to future events could easily be captured by one or more factors.

To make the comparison to the CAPM as concrete as possible, assume that factor analysis identifies two priced factors $\tilde{F}_1$ and $\tilde{F}_2$ that are significant. Consider the following scenarios:

1. The CAPM (with a Risk Free Asset) is True and We Use a Good Proxy for the Market. It is possible that $r_m = a + b_1 \tilde{F}_1 + b_2 \tilde{F}_2$, that is, both factors are correlated with the market but the market portfolio is a sufficient statistic. This is possible in an APT model with theoretically specified factors. In factor analysis, however, in this case only one factor should be picked which would be perfectly correlated with the market return.

2. The CAPM (with a Risk Free Asset) is True but We Have a Poor Proxy for the Market.
A market factor may work, but other factors should help also to capture the true market portfolio. Finding additional factors uncorrelated with the market proxy only proves that the proxy is bad; it does not disprove the CAPM.

3. The CAPM is Not True but the APT Holds. A second factor in addition to the market portfolio may become significant in a variety of cases. For instance if returns are lognormally distributed. Then it is not just portfolio variance that matters and the other factor may capture the skewness or other higher moments of the lognormal distribution. Similarly, if we have nonmarketable assets, foreign trade, non-homothethic utility, or dynamic hedging effects. As we saw for most of these cases, the resulting extension of the CAPM implies additional betas and these then would be picked up in the factor analysis.

(g) Applications and Exercises

1. Suppose that an APT model holds with only one factor. Does this factor have to be the market portfolio? Explain.

2. Consider a 2-factor APT without idiosyncratic risk. Assume that one asset lies above the equilibrium asset pricing plane. Explain how an arbitrage portfolio can be constructed in this case.

5. THE FAMA-FRENCH THREE FACTOR MODEL

(a) Description of the Empirical Model and Results

Fama and French (1992) shocked the finance profession by showing that, with recent data, market β can no longer account for the cross-sectional variation in stock returns. For the period 1963-1990, beta seems to play no role in explaining the average returns on NYSE, AMEX, and NASDAQ. Fama and French obtain these results in a set-up that is related to the Fama-MacBeth methodology, but with some significant differences. First, portfolios are not the 20 portfolios pre-sorted by beta but are the 100 portfolios obtained by sorting first into size deciles and then, within each size decile, into beta deciles. Fama and French claim that this way of sorting is crucial since size and market beta are strongly correlated. Previous studies may have indicated a role for beta that, instead, should have been attributed to the size variable. Second, betas are calculated based on all post-sorting data (that is from July 1963 to December 1990); thus the beta of each of the 100 portfolios does not change through time. Third, betas are estimated as the sum of the slopes in the regression of the portfolio return on current and prior month market returns (that is, both market return and lagged market return are included as explanatory variables in the first-pass regressions). According to Dimson (1979) this adjusts for nonsynchronous trading and may be most relevant for smaller firms. Fourth, cross-sectional (second-pass) regressions are conducted on individual stocks rather than portfolios; with each individual stock assigned the beta of the size-beta sorted portfolio it currently belong to. This is done for efficiency reasons because other variables (the "x" variables) in the second-pass regression can be estimated reliably for individual firms.

The deviations from the standard Fama-MacBeth approach seem reasonable, especially from the perspective of pointing out the failure of beta in explaining the cross-section of stock returns: the use of future information in estimating betas, if anything, will provide a bias in support of beta. While one may argue that time variation in betas provides a problem, Fama and French state that results stand up to robustness checks when 60 months instead of the full sample are used in estimating beta. In addition, Chan and Chen (1988) show that full-period beta estimates can work well even if betas vary through time.

The result of the second-pass regression for all securities explaining the cross-section of monthly returns, when only beta (as obtained using the aforementioned deviations from the Fama-MacBeth approach) is used as the right-hand variable is a slope of 0.15% with a t-statistic of 0.46; the death blow to the CAPM. To make matters worse, Fama and

French (1992) show that both size, measured as the log of market equity–price per share times shares of common stock outstanding–as well as book-to-market ratios, measured as the log of book value–accounting value of common stock–minus the log of market equity, are statistically significant in explaining the cross-sectional returns of the, on average, 2267 stocks in the sample; the slope of the size variable is significantly negative and the slope of the book-to-market variable is significantly positive.

Fama and French (1993) expand on their 1992 paper by considering corporate and government bonds in addition to common stock as part of the asset returns to be explained. Accordingly, they also expand the set of explanatory factors under consideration by adding a term structure and a default premium variable; factors known to have explanatory power for the cross-section of bond returns. Furthermore, instead of the Fama-MacBeth two-pass regression approach they consider the single pass time-series regression approach in which the key test is whether the mean of the unexplained variation in the asset pricing model, the "alpha", is significantly positive.

Fama and French now, in effect, use five factors: beta, the book-to-market ratio, size, the default premium, and the term structure variable to explain the cross-section of stock and bond returns. Bonds are grouped into seven portfolios: five corporate groups sorted by their Moody's credit rating and government bonds split by maturity–more than five years and five or less years to maturity. They find that bonds are explained well by "their own" term structure and default premium factors (R^2 of around 0.90 or higher except for the short-term government bonds, R^2 of 0.79, and the lowest grade corporate bonds, R^2 of 0.49). In addition, the remaining factors have little explanatory power for bonds. Similarly, stocks are well explained by "their" three factors: market beta, size, and book-to-market ratio. These results are a little surprising, in the sense that standard asset pricing theory suggests that any asset should be priced by the same factors–those that indicate the asset's impact on the wealth or consumption risk of the representative investor. Apparently, though, stocks have small loadings on the two "bond" factors and bonds have small loadings on the three "stock" factors. (According to the CAPM, of course, all average asset returns should be explained by the same factor, market risk.) There seems to be some degree of interaction, however, between the stock and bond markets because, once the correlation of market beta with the default premium and the terms structure is removed, the default premium and term structure premium do have a significant effect in explaining stock returns.

We next turn to the explanation of stock returns. Fama and French (1996) extend their 1993 study to focus solely on explaining the cross-sectional pattern of stock returns and to investigate how the three stock factors absorb previously discovered anomalies (patterns in returns not explained by the CAPM; related to the explanatory power of size, price-earnings ratios, cash-flow to price ratios, leverage ratios, long-term past return, short-term past return, and sales growth). In the following we provide a detailed description of this "three-factor model" for stock returns as given in Fama and French (1993, 1996).

Based on the empirical findings in Fama and French (1992), Fama and French posit the following factor model as applies to any portfolio i :

$$\mu_i - r_f = b_i\,(\mu_m - r_f) + s_i\,\mu_{SMB} + h_i\,\mu_{HML}\,, \quad (1)$$

where $\mu_m - r_f$ indicates the market risk premium–the mean return on the market factor, μ_{SMB} equals the mean return on a "size" factor, and μ_{HML} represents the mean return on a "book-to-market" factor. The size factor is the expected return on a zero-investment portfolio that is long a portfolio of small firms and is short a portfolio of big firms (*SMB* stands for small minus big); the book-to-market factor is a zero-investment portfolio that is long a portfolio of firms with high book-to-market ratios and short a portfolio of firms with low book-to-market ratios (*HML* stands for high minus low). To generate the factor values, Fama and French create six portfolios by splitting the firms in two size groups and in three book-to-market groups. Why they create six portfolios instead of nine or four is not clear; Fama and French admit that the choice is arbitrary but that they have not searched over alternatives. Typically, however, researchers cannot get away with arbitrary choices such as these. The high book-to-market portfolio *H* consists of the 30% stocks with the highest book-to-market ratios in a given year; the low book-to-market portfolio *L* consists of the 30% stocks with the lowest book-to-market ratios; the portfolio *M* contains the remaining 40% of stocks which are not considered in generating the *HML* factor. The *HML* factor returns are obtained by subtracting the return on *L* from the return on *H*. Similarly, a small firm portfolio *S* is formed containing the smaller firms (more than 50% to guarantee that this portfolio does not consist of AMEX and NASDAQ firms alone) and a big firm portfolio *B* is formed containing the larger firms (the remaining firms). The *SMB* factor returns are obtained by subtracting the return on *S* from the return on *B*. Clearly, in view of previous empirical evidence of a size effect, the mean return on the zero-investment size portfolio, μ_{SMB}, is expected to be positive, and from the evidence of abnormal positive returns of "value" firms–those with high book-to-market ratios–compared to "growth" firms–those with low book-to-market ratios–the book-to-market mean return, μ_{HML}, is expected to be positive as well.

The slopes b_i, s_i, and h_i (the factor loadings) are estimated from a time-series regression, which also produces the "mis-pricing" residuals α_i (the alphas):

$$(2) \qquad r_i - r_f = \alpha_i + b_i (r_m - r_f) + s_i r_{SMB} + h_i r_{HML} + \varepsilon_i .$$

The formal time series test of any factor model is to check if the intercept α_i for all *i* is jointly significantly different from zero. The formal test statistic for this test is an F-statistic as derived by Gibbons, Ross, and Shanken (1989).

The portfolios *i* are chosen by splitting the sample in size and book-to-market quintiles, generating 25 portfolios. Then 25 time-series regressions of the form of equation (2) are run, each for the 366 months from July 1963 to December 1993. Each of these regressions have an R^2 above 0.80 and typically above 0.90. Nevertheless, the Gibbons-Ross-Shanken test convincingly rejects the null-hypothesis that the alphas are jointly equal to zero (although the rejection is more convincing for the CAPM case for which the s_i and h_i coefficients are restricted to equal zero). Fama and French argue that the rejection is expected since the high power of the test will pick up even very small deviations from zero that should occur in any model since, by definition, a model is not reality. They further argue that the average size of the alphas across the 25 time series is economically very small, around 9 basis points per month. Fama and French also discuss that the size and book-to-market factors are useful in explaining cross-sectional differences in returns (as we know from their 1992 paper); the market factor is not useful for that purpose (as we also know from their 1992 paper) but explains most of the discrepancy between average stock returns and the average one-month T-Bill rate (used here as the risk free rate). Hence, market excess return is included here as a factor for that reason.

Numerically, the returns of the 25 portfolios vary from a monthly average of 0.32% to 1.05% which is a huge difference for portfolios consisting of an average number of around 100 firms (varying, though, from 24 in the biggest value firms portfolio to over 500 in the smallest value firms portfolio). The average market risk premium, that is

$\mu_m - r_f$, is 0.43% per month; the average size risk premium, μ_{SMB}, is 0.27% a month; and the average book-to-market risk premium, μ_{HML}, is 0.40%. Only the latter is significantly different from zero. While r_{HML} and r_{SMB} are basically uncorrelated, the correlations of r_{HML} and r_{SMB} with $r_m - r_f$ are, respectively, -0.38 and 0.32. The slopes on the three factors are strongly significant in all cases. The t-statistics for the b_i in all 25 cases exceed 38.6. For the s_i the t-statistics generally exceed 10.0 except for the biggest firms, which have small loadings on the size factor. For the h_i the t-statistics vary from strongly negative -10.0 or lower to strongly positive, larger than 20.0 as we go from low to high book-to-market portfolios; clearly the high book-to-market firms have high loadings on the book-to-market factor. The market betas (the b_i) differ very little, all being close to one, across the 25 portfolios (by design almost, since the portfolios were only sorted by size and book-to-market ratio) so that, for these portfolios, market beta has little importance in explaining the cross-section of the 25 portfolio returns. The average market risk premium thus explains the level of the average returns relative to the one-month T-Bill rate. The size betas (the s_i) vary from around -0.10 for big firms to 1.30 for large firms. Given the μ_{SMB} of 0.27% a month, this explains an annual premium of around 4.5% for small firms over large firms. The book-to-market betas (the h_i) vary from around -0.40 for low book-to-market firms to 0.70 for high book-to-market firms. Given the μ_{HML} of 0.40% a month, this explains an annual premium of around 5.3% for value stocks compared to growth stocks. To some these premiums are a little too high for murky risk factors to be rational, but they are in the same ballpark as the equity premium of around 5.2% in these data.

(b) Discussion

We can think of the three-factor model as an example of an APT model with three factors (given that a riskless asset exists and excess returns are used as left-hand-side variables so that the intercept must be zero), or as an example of an Intertemporal CAPM model such as we will discuss later. Unfortunately, the interpretation of the factors as risk factors is not straightforward. Fama and French admit that the factors were generated in an ad hoc fashion but attempt to provide some explanation for their importance. Clearly, the market excess return needs no motivation as a factor. The book-to-market factor, according to Fama and French, captures the notion of "distress." Firms with high book-to-market ratios tend to have low market prices due to previously low earnings; as a result they are more likely to be near bankruptcy, or at least financially distressed. Therefore, they may not improve as the market improves, and provide an independent source of risk. Smaller firms also may provide a risk that differs from general market risk. For one, there may be more of an adverse selection risk to outside investors as the shares are more closely held, and they may be less liquid which may impose a liquidity premium such as those in term structure models. Fama and French do not propose these latter possibilities, however, but prefer to say that, as to size risk, they are not sure how it can be explained.

In spite of the theoretical drawbacks, the three-factor model does explain most of the standard anomalies discovered since the advent of the CAPM in the sixties. First, note that a book-to-market ratio is inversely related to the market price of a particular stock. However, the same can be said about earnings-to-price ratios, dividend-to-price ratios, and cashflow-to-price ratios. These variables in past empirical work have all been found to explain cross-sectional differences in stock returns. It is not surprising that, once one of these variables is included, the others lose their predictive power as Fama and French find. Similarly, the existence of mean reversion that, as DeBondt and Thaler (1985) show, leads to profitability of contrarian strategies–selling stocks that have done well over the last three to five years and buying those that have done poorly over that time–implies that "high" market prices lead to subsequent lower average returns, is again captured by the book-to-market factor as low book-to-market ratios imply both lower average returns as well as tending to imply "high" market prices. Sales growth is another variable that may be closely correlated with book-to-market ratios as higher sales growth typically implies lower market prices. Leverage (which is found to have a positive effect on average stock returns when equity value is measured in market terms and a negative effect when equity value is measured in book terms), is also correlated with book-to-market ratios: negatively when market equity values are used and positively when book values are used. Lastly, the size effect of Banz (1981) is obviously captured by the size variable in the three-factor model. However, Fama and French find that the "momentum" effect documented by Jegadeesh and Titman (1993)–stocks with high returns in the recent past tend to maintain that pattern in the nearby future–cannot be subsumed by the three-factor model. Their comment in 1996 was that one possibility is that this momentum anomaly will disappear once it is investigated more closely. However, at this time it seems that evidence for a momentum anomaly continues to mount.

What does the three-factor model tell us about asset pricing in general? As always there are three types of explanations. First, the three-factor model is not "real"; it is the result of data mining which means that it should not be useful in the future. Or it could be due to a measurement problem such as that related to proper measurement of the market return. In this case the CAPM could still be true and the two non-market factors just help to provide a better estimate of the true market return. The three-factor model then should be useful in the future, for instance in making proper risk corrections. Second, the model is "real" and rational explanations along the lines of the APT or ICAPM explain why the non-market factors are priced. Problem is of course that we do not yet have a good idea of what sort of risk is being priced. Again, though, the three-factor model now should be valuable in making risk corrections. Third, non-rational or quasi-rational explanations hold to explain the deviations from the CAPM. In this case, the factor returns really present excess profit opportunities that can be exploited if they persist over time. Persistence if, of course, a big if. Behavioral finance maintains that there are behavioral biases in the actions of investors that are systematic and thus are likely to persist. The three-factor model now would be inappropriate for making risk corrections.

Fama and French point out the different applications for which their three-factor model would dominate the CAPM which is, currently, still popular with practitioners. First, in event studies, especially when these are conducted over a large time interval, excess returns should be corrected for risk using the three-factor model. Second, in evaluating portfolio managers or mutual funds, the realized returns should be adjusted for the risk based on the three-factor model and the resulting return should be compared to the alphas derived in their study. Third, portfolio choice should be guided by the fact that proper hedging should occur with respect to all three of the factors. Fourth, provided that the factor loadings of individual investment projects can be estimated accurately (regarding which Fama and French express some reservations), the three-factor model can be used for calculating the cost of capital to be used in capital budgeting decisions.

(c) Applications and Exercises

1. Perform a simple test of the Fama-French three factor model. The question is: how well does this model explain the cross-section of *average* returns of portfolios sorted by size and value characteristics?

   Seven computer files are available from me. Two text files with data obtained from Kenneth French's web site at MIT: monthly returns on 25 portfolios sorted by book-to-market ratio and size; and monthly values for the size factor (SMB), the value factor (HML), the market factor (Market minus riskfree return), and the riskfree rate. The time period is from July 1926 to June 2000. These two files are sufficient to complete the assignment.

The five remaining files are a data work file and four (pretty clumsy) batch programs that may greatly simplify your assignment and allow you to run the assignment below in EViews (if you so choose). The batch files: (1) obtain mean values for the 25 portfolio returns (meanslij.pgm), (2) run a time series "first-pass" regression to obtain full-sample values for the three Fama-French betas of each of the 25 portfolios (regs.pgm), (3) organize the data for a second-pass regression (coeff.pgm), and (4) run the cross-sectional "second-pass" regression for the mean returns of the 25 portfolios (final.pgm).

Open *(Open→ Work File)* the work file in EViews (available on the I-drive), then Open *(Open→ Program)* each program and *Run* the four batch files in turn to get the desired results. Or import the text files into, say SAS, and write your own version of these batch files.

Questions

(a) Interpret the results from the second pass.

(b) Obtain analogous results for the CAPM version when only the Market beta is used. [Just appropriately edit the EViews batch files for the easiest way to get these results].

(c) Are the results generally consistent with those of Fama and French (1992)?

(d) Compare the approach used here to that in Fama and MacBeth (1973) and that in Fama and French (1992) and discuss the differences.

(e) Could you directly use the factor values given in the original data file to make risk corrections for other groups of portfolios? Explain.

6. OTHER VARIANTS OF THE CAPM

Various attempts have been made to improve on the CAPM by dropping one or more of the less desirable assumptions. Here we discuss some of these without much detail. Reason for this in part is that these extensions, so far, have not proven to be very successful both empirically and theoretically.

(a) The Partial Variance Approach

The idea is that the variance of above average returns realizations is irrelevant for the consideration of risk. As such, only the variance below a particular threshold is calculated. The threshold is typically set equal to the risk free return. Harlow and Rao (1989) show in general that in this scenario a one-beta CAPM obtains, where however the beta cannot be estimated by standard regression methods. They instead estimate beta by (among other changes) separating the market return variable into two separate variables: one for when the market return is above the threshold (this variable is zero whenever market return is *below* the threshold) and one for when the market return is below the threshold (this variable is zero whenever market return is *above* the threshold). The coefficient on the variable for when the market return is below the threshold becomes the appropriate beta.

(b) The Three-Moment CAPM

If we have a better approximation of preferences than quadratic utility then we no longer need to assume ellipticality of returns. This is possible by considering a three-moment or four-moment CAPM where apart from mean and variance also the skewness and kurtosis of the portfolio return matter.

Kraus and Litzenberger (1976) developed a three moment CAPM by considering a third-order Taylor approximation for the utility function of an investor. Here we, equivalently, take a second-order approximation of marginal utility around the initial level of wealth:

$$(1) \qquad u'(w) \approx u'(\bar{w}) + u''(\bar{w})(w - \bar{w}) + [u'''(\bar{w})/2](w - \bar{w})^2$$

Assuming, as is standard, positive marginal utility and risk aversion, the reasonable restriction of DARA (Decreasing Absolute Risk Aversion) preferences (which avoids that risky assets are inferior goods) implies a preference for higher skewness: $sgn[d(-u''/u')/dw] = sgn[-u'''u' + (u'')^2] < 0$ requires that $u''' > 0$.[5]

To derive a three-moment CAPM a simplifying assumption (not made by Kraus and Litzenberger) is that a *representative investor* exists. We know then that $w - \bar{w} = r_m \bar{w}$; that is, the market return is the return on wealth for the representative investor. Equation (1) can then be written as:

$$(2) \qquad u'(w) \approx g_0 - g_1 r_m + g_2 r_m^2 \quad , \quad g_j > 0 \ \text{ for all } j.$$

Given a representative investor and without assuming ellipticality of returns, Chapter III, equation (2.13) gives:

$$(3) \qquad E[u'(w)(r_i - r_f)] = 0, \quad \text{for all assets } i.$$

Using the definition of covariance, equation (3) yields:

$$(4) \qquad \mu_i - r_f = -Cov[u'(w), r_i]/E[u'(w)], \quad \text{for all assets } i.$$

Substituting equation (2) into (4) yields:

[5] For expected utility we have $E[u(w)] \approx u(\bar{w}) + [\bar{w}^2 u''(\bar{w})/2]\,\sigma_m^2 + [\bar{w}^3 u'''(\bar{w})/6]\,s_m^3$, where $s_m^3 \equiv E(r_m - \mu_m)^3$ indicates the third (central) moment, the skewness, of the market return distribution. Positive skewness (longer tail to the right of the distribution) raises expected utility as people like upward potential with little downside risk, keeping mean and variance constant.

(5) $$\mu_i - r_f = h_1 \sigma_{im} - h_2 s_{imm} , \quad \text{for all assets } i \text{ and with } h_1, h_2 > 0 ,$$

where $s_{imm} \equiv Cov(r_i, r_m^2)$ indicates the co-skewness between asset i's return and the market return. Kraus and Litzenberger define a beta and a "gamma" as follows: $\beta_i \equiv \sigma_{im}/\sigma_m^2$, $\gamma_i = s_{imm}/s_m^3$ [where $s_m^3 \equiv E(r_m - \mu_m)^3$]. This is not essential in developing and testing the three-moment CAPM and we will skip that step; it should however be clear that equation (5) represents essentially a two-beta CAPM.

The approach we employed previously in finding the asset pricing equation is to apply equation (5) to two "benchmark" assets, the market portfolio and one "other" asset. Since we have no criterion here for choosing the "other" asset, an alternative approach is more useful. We can always write the following tautology:

(6) $$r_i - r_f = c_{0i} + c_{1i}(r_m - r_f) + c_{2i}(r_m - \mu_m)^2 + \varepsilon_i ,$$

where $E(\varepsilon_i) = Cov(\varepsilon_i, r_m) = Cov[\varepsilon_i, (r_m - \mu_m)^2] = 0$. Subtracting from both sides of equation (6) their expected values, then multiplying both sides by $r_m - \mu_m$ and then taking expected values gives:

(7) $$\sigma_{im} = c_{1i}\sigma_m^2 + c_{2i}s_m^3 .$$

The same process but multiplying both sides by $(r_m - \mu_m)^2$ gives:

(8) $$s_{imm} = c_{1i}s_m^3 + c_{2i}k_m^4 ,$$

where $k_m^4 \equiv E(r_m - \mu_m)^4$ represents the kurtosis, the fourth (central) moment of the market return distribution.

It is now straightforward to test the three-moment CAPM using a two-pass regression approach. First, empirically obtain the first four central moments of the market return distribution; then find the $\hat{c}_{ji}$ from regression equation (6) for all assets (or portfolios) to execute the first pass. This will provide estimates $\hat{\sigma}_{im}$ and $\hat{s}_{imm}$ for all assets. Second, employ the first-pass estimates to test equation (9) and find estimates $\hat{h}_1$, $\hat{h}_2$ which should be significantly positive in all cases (even when the realized market excess return is negative). Kraus and Litzenberger's results support the three-factor CAPM as they find significantly positive estimates of $\hat{h}_1$, $\hat{h}_2$ in the second pass regression. Note that their approach is very similar to the approach outlined here only because of the fact that market returns have positive skewness; otherwise the $\gamma_i = s_{imm}/s_m^3$ which they use and s_{imm} which we use would have opposite signs.

(c) The Four-Moment CAPM

A straightforward extension of the above approach yields the four-moment CAPM. This extension of the CAPM is developed by Fang and Lai (1997) and Dittmar (1999). It is relevant because of the well-known observation of kurtosis, thick tails, in stock returns. In addition, there is some theoretical support for the relevance of the fourth moment in the utility function. Kimball develops the concept of Decreasing Absolute Prudence. Together with non-satiation, risk aversion and decreasing absolute risk aversion, decreasing absolute prudence implies that $u^{////}(w) < 0$. Thus, individual investors are presumed to be averse to kurtosis–for given variance, more extreme outcomes are disliked.

In the four-moment CAPM three betas become relevant, related to co-variance, co-skewness, and co-kurtosis (common sensitivity to extreme outcomes) of an asset's return with the market return. Dittmar reports good results in a conditional version of this model relative to the Fama-French three factor model.

(d) Applications and Exercises

1. Employing the approach in section 6(b), explicitly derive the four-moment CAPM and discuss the differences relative to the three-moment CAPM.