Factor Models in Portfolio and Asset Pricing Theory

Gregory Connor*and Robert Korajczyk†‡

May 29, 2008

Abstract

The foundation of modern portfolio theory is the mean-variance portfolio selection approach of Markowitz (1952, 1959). We discuss the role of factor models in implementing portfolio selection, defining the nature of systematic risk, and estimating the premium for risk bearing.

*Department of Finance, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: g.connor@lse.ac.uk.

†Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2001, USA. E-mail address: r-korajczyk@kellogg.northwestern.edu. Korajczyk wishes to acknowledge the financial support of the Zell Center for Risk Research and the Jerome Kenney Fund.

‡Forthcoming in *Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques*, edited by John Guerard, London: Springer.

1 Portfolio Selection

The mean-variance approach of Markowitz (1952, 1959) is an essential underpinning of modern portfolio theory and asset pricing theory. Rational investors who care about expected returns and the variance of their portfolio will hold, subject to any constraints they face, "efficient" portfolios. That is, they hold portfolios with the highest expected return per unit of risk (either variance or standard deviation). Important insights from portfolio selection are that investors rationally diversify (depending on the constraints they face) and that the volatility of an asset (variance or standard deviation) is not, in general, a good indicator of its contribution to the risk of a portfolio of many assets.

At the risk of oversimplifying the analysis of Markowitz, consider the case in which investors have no constraints on short sales.[1] Let ω be a vector of asset positions, with ω_i denoting the fraction of the investor's wealth in asset i (where $i = 1, 2, ...n$). Positive values of ω_i correspond to long positions and negative values to short positions. We require that 100% of the investor's wealth is allocated to assets, that is, $\omega' \iota = 1$, where ι is an $n-$vector of ones. Letting μ be the $n-$vector of expected returns on assets and V be the $n \times n$ covariance matrix of asset returns, the portfolio's expected return (μ_p), Variance (σ_p^2), and standard deviation (σ_p) are:

$$\mu_p = \omega' \mu$$

$$\sigma_p^2 = \omega' V \omega$$

$$\sigma_p = \sqrt{\omega' V \omega}$$

Let $A = \iota' V^{-1} \mu$, $B = \mu' V^{-1} \mu, C = \iota' V^{-1} \iota$, and $D = BC - A^2$. The portfolio weights that minimize variance, for a given expected return, μ_p, are given by (see Huang and Litzenberger (1988, section 3.8) or Cochrane (2001, section 5.2)):

$$\omega_{MV} = a + b\mu_p$$

where

$$\begin{aligned} a &= \frac{1}{D}[BV^{-1}\iota - AV^{-1}\mu] \\ b &= \frac{1}{D}[CV^{-1}\mu - AV^{-1}\iota]. \end{aligned}$$

Even in this simple case, the solution for minimum-variance portfolio weights requires the inversion of the covariance matrix, V. In practice, of course, we do not observe V, but must estimate it. When there are many assets the solution may be infeasible, due to the computational burden, or unreliable, due

[1] Markowitz considered many more complicated problems than the one considered here.

to the estimation error in $\widehat{V}$. One approach is to make simplifying assumptions about the structure of V to make the solution easier. One way to impose such restrictions on the covariance matrix is to assume a factor model as the return generating process.

2 Basic Definition of a Factor Model

Factor models of security returns decompose the random return on each of a cross-section of assets into factor-related and asset-specific returns. Letting r denote the vector of random returns on n assets, and assuming k factors, a factor decomposition has the form:

$$r = a + Bf + \varepsilon \tag{1}$$

where B is a $n \times k$-matrix of factor betas, f is a random $k-$vector of factor returns, and ε is an $n-$vector of asset-specific returns. The $n-$vector of coefficients a is set so that $E[\varepsilon] = 0$. By defining B as the least squares projection $B = cov(r, f)V_f^{-1}$, it follows that $cov(f, \varepsilon) = 0^{k \times n}$.

The factor decomposition (1) puts no empirical restrictions on returns beyond requiring that the means and variances of r and f exist. So in this sense it is empty of empirical content. To add empirical structure it is commonly assumed that the asset-specific returns ε are cross-sectionally uncorrelated, $E[\varepsilon\varepsilon'] = D$ where D is a diagonal matrix. This implies that the covariance matrix of returns can be written as the sum of a matrix of rank k and a diagonal matrix:

$$V = cov(r, r') = BV_f B' + D. \tag{2}$$

This is called a strict factor model. Without loss of generality one can assume that $V_f = cov(f, f')$ has rank k, since otherwise one of the factors can be removed (giving a $k-1$ factor model) without affecting the fit of the model.

If the vector of factors, f, is a specified set of random variables (e.g., the returns on a set of market indices), we will call the model a pre-specified, or macroeconomic, factor model. If the vector of factors, f, is unknown, and must be estimated, we will call the model a statistical factor model. If the matrix of factor loadings, B, is known, but f must be estimated, we will call the model a characteristic factor model (Connor (1995)). There are also applications which involve hybrid factor models: ones which have some pre-specified factors and/or loadings and some statistical factors (Stroyny (2005)).

Estimation of the unrestricted covariance matrix of n securities, V, requires the estimation of $n \times (n+1)/2$ distinct elements. The single index, or diagonal, model of Sharpe (1963) is a single-factor pre-specified strict factor model. It assumes that all of the common elements of returns were due to assets' relations with the market index. Thus, only $(3 \times n) + 1$ parameters needed to be estimated: n elements of B, or "betas" relative to the index, n unique variances, n intercept terms, and the index variance, σ_f^2. This approach reduces much of the noise in the estimate of Σ. In practice, the single index, diagonal model does not describe all of the common movements across assets (i.e., the residual matrix is not diagonal) so there seems to be some additional benefit from using a multifactor model. With k factors there are still only $n \times (k+2) + k \times (k+1)/2$ parameters to estimate ($n \times k$ betas, n intercepts or means, n unique variances, and $k \times (k+1)/2$ elements of the factor covariance matrix). An alternative approach is to continue with a smaller set of factors, but place other restrictions on the covariance matrix of ε. Some early studies in this area are Farrar (1962), King (1966), Cohen and Pogue (1967), and Elton and Gruber (1973).

A strict factor model aids in computing optimal portfolio weights since

$$V^{-1} = D^{-1} - D^{-1}BV_f(V_f + V_fB'D^{-1}BV_f)^{-1}V_fB'D^{-1}$$

(see Muirhead (1982, p. 580)). The strict factor model only requires the inversion of an $n \times n$ diagonal matrix, D, and a $k \times k$ matrix, $V_f + V_fB'D^{-1}BV_f$. Since k is typically very small, relative to the number of assets, n, these inverses are easy to calculate and are likely to be more stable than the inverse of an unrestricted $n \times n$ covariance matrix. The factor model trades off the potential bias that the residual covariance matrix may not be diagonal with the added precision of the model estimates due to the higher numbers of observations per parameter.

3 Statistical Factor Models

We differentiate between characteristic-based, macroeconomic, and statistical factor models. In a characteristic-based model the factor betas of asset are tied to observable characteristics of the securities, such as company size or the book-to-price ratio, or the industry categories to which each security belongs. In macroeconomic factor models, the factors are linked to the innovations in observable economic time series such as inflation and unemployment. In a statistical factor model, neither factors nor betas are tied to any external data sources and the model is identified from the covariances of asset returns alone.[2]

The convenient rotation $E[ff'] = I$ allows us to write the strict factor model (1) as:

$$cov(r, r') = BB' + D. \tag{3}$$

Assuming that the cross-section of return is multivariate normal and i.i.d. through time, the sample covariance matrix $\widehat{cov}(r, r')$ has a Wishart distribution. Imposing the strict factor model assumption (3) on the true covariance matrix it is possible to estimate the set of parameters B, D by maximum likelihood. This maximum likelihood problem requires high-dimensional nonlinear maximization: there are $nK + n$ parameters to estimate in B, D. There is also an inequality constraint on the maximization problem: the diagonal elements of D must be nonnegative, since they represent variances. The solution to the maximum likelihood problem yields estimates of B and D which correspond to the systematic and unsystematic risk measures. It is often the case that estimates

[2] See Connor (1995) and Connor and Korajczyk (2007).

of the time series of factors, f, are of interest. These are called factor scores in the statistical literature and can be obtained through cross-sectional GLS regressions of r on B

$$\widehat{f}_t = (\widehat{B}\widehat{D}^{-1}\widehat{B})^{-1}\widehat{B}\widehat{D}^{-1}r_t.$$

See Basilevsky (1994) for a review of the various iterative algorithms which can be used to numerically solve the maximum likelihood factor analysis problem and estimate the factor scores. See Roll and Ross (1980) for an empirical application to equity returns data.

The first k eigenvectors of the return covariance matrix scaled by the square roots of their respective eigenvalues are called the k *principal components* of the covariance matrix. A restrictive version of the strict factor model is the *scalar factor model*, given by (2) plus the scalar matrix condition $D = \sigma_\varepsilon^2 I$. Under the assumption of a scalar factor model, the maximum likelihood problem simplifies, and the principal components are the maximum likelihood estimates of the factor beta matrix B (the arbitrary choice of rotation is slightly different in this case). This provides a quick and simple alternative to maximum likelihood factor analysis, under the restrictive assumption $D = \sigma_\varepsilon^2 I$.

An important (and often troublesome) feature of statistical factor models is their rotational indeterminacy. Let L denote any nonsingular $k \times k$-matrix and consider the set of factors $f^* = Lf$ and factor betas $B^* = L^{-1}B$. Note that f^*, B^* can be used in place of f, B since only their matrix product affects returns and the linear "rotation" L disappears from this product. This means that factors f and associated factor betas B are only defined up to a $k \times k$ linear transformation. In order to empirically identify the factor model one can set the covariance matrix of the factors equal to an identity matrix, $E[ff'] = I_k$, without loss of generality.

3.1 Approximate Factor Models

The assumption that returns obey a strict factor model is easily rejected. In practice, for most reasonable values of k there will at least some discernible positive correlations between the asset–specific returns of at least some assets. An *approximate factor model* (originally developed by Chamberlain and Rothschild (1983)) weakens the strict factor model of exactly zero correlations between all asset specific returns. Instead it assumes that the there is a large number of assets n and the proportion of the correlations which are nonnegligibly different from zero is close to zero. This condition is formalized as a bound on the eigenvalues of the asset-specific return covariance matrix:

$$\lim_{n\to\infty} \max eigval[cov(\varepsilon, \varepsilon')] < c$$

for some fixed $c < \infty$. Crucially, this condition implies that asset-specific returns are *diversifiable risk* in the sense that any well-spread portfolio w will have asset-specific variance near zero:

$$\lim_{n\to\infty} w' cov(\varepsilon, \varepsilon')w = 0 \text{ for any } w \text{ such that } \lim_{n\to\infty} w'w = 0. \tag{4}$$

Note that an approximate factor model uses a "large n" modeling approach: the restrictions on the covariance matrix need only hold approximately as the number of assets n grows large.

Letting $V = cov(\varepsilon, \varepsilon')$ which is no longer diagonal, and choosing the rotation so that $cov(f, f') = I$ we can write the covariance matrix of returns as:

$$cov(r, r') = BB' + V.$$

In addition to (4) it is appropriate to impose the condition that $\lim_{n\to\infty} \min BB' = \infty$. This ensures that each of the k factors represents a pervasive source of risk in the cross-section of returns.

3.2 Asymptotic Principal Components

The maximum likelihood method of factor model estimation relies on a strict factor model assumption and a time-series sample which is large relative to the number of assets in the cross-section. Standard principal components requires the even stronger condition of a scalar factor model. Neither method is well-configured for asset returns where the cross-section tends to be very large. Connor and Korajczyk (1986) develop an alternative method called asymptotic principal components, building on the approximate factor model theory of Chamberlain and Rothschild (1983). Connor and Korajczyk analyze the eigenvector decomposition of the $T \times T$ cross product matrix of returns rather than of the $n \times n$ covariance matrix of returns. They show that given a large cross-section, the first k eigenvectors of this cross-product matrix provide consistent estimates of the $k \times T$ matrix of factor returns. Stock and Watson (2002) extend the theory to allow both large time series and large cross-sectional samples, time varying factor betas, and provide a quasi-maximum likelihood interpretation of the technique. Bai (2003) analyzes the large-sample distributions of the factor returns and factor beta matrix estimates in a generalized version of this approach.

4 Macroeconomic Factor Models

The rotational indeterminacy in statistical factor models is unsatisfying for the application of factor models to many research problems. Statistical factor models do not allow the analyst to assign meaningful labels to the factors and betas; one can identify the k pervasive risks in the cross-section of returns, but not what these risks represent in terms of economic and financial theory.

One approach to making the factor decomposition more interpretable is to rotate the statistical factors so that the rotated factors are maximally correlated with pre-specified macroeconomic factors. If f_t is a k-vector of statistical factors and m_t is a k-vector of macroeconomic innovations we can regress the macroeconomic factors on the statistical factors

$$m_t = \Pi f_t + \eta_t.$$

As long as Π has rank k, the span of the rotated factors, Πf_t, is the span of the original statistical factors, f_t. However the rotated factors can now be interpreted as the return factors that are correlated with the specified macroeconomic series. With this rotation the new factors are no longer orthogonal, in general. This approach is described in Connor and Korajczyk (1991).

Alternatively, one can work with the pre-specified macroeconomic series directly. Chan, Chen, and Hsieh (1985) and Chen, Roll and Ross (1986) develop a macroeconomic factor model in which the factor innovations f are observed directly (using innovations in economic time series) and the factor betas are estimated via time-series regression of each asset's return on the time-series of factors. They begin with the standard description of the current price of each asset, p_{it}, as the present discounted value of its expected cash flows:

$$p_{it} = \sum_{s=1}^{\infty} \frac{E[c_{it}]}{(1+\rho_{st})^s}$$

where ρ_{st} is the discount rate at time t for expected cash flows at time $t+s$. Chen, Roll and Ross note that the common factors in returns must be variables which cause pervasive shocks to expected cash flows $E[c_{it}]$ and/or risk-adjusted discount rates ρ_{st}. They propose inflation, interest rate, and business-cycle related variates to capture these common factors. Shanken and Weinstein (2006) find that empirically the model lacks robustness in that small changes in the included factors or the sample period have large effects on the estimates. Connor (1995) argues that although macroeconomic factors models are theoretically attractive since they provide a deeper explanation for return comovement than statistical factor models, their empirical fit is substantially weaker than statistical and characteristic-based models. Vassalou (2003) argues on the other hand that the ability of the Fama-French model (see below) to explain the cross-section of mean returns can be attributed to the fact that Fama-French factors provide good proxies for macroeconomic factors.

5 Characteristic-based Factor Models

A surprisingly powerful method for factor modeling of security returns is the characteristic-based factor model. Rosenberg (1974) was the first to suggest that suitably scaled versions of standard accounting ratios (book-to-price ratio, market value of equity) could serve as factor betas. Using these predefined betas, he estimates the factor realizations f_t by cross-sectional regression of time-t asset returns on the pre-defined matrix of betas.

In a series of very influential papers, Fama and French (1992, 1993, 1996) propose a two-stage method for estimating characteristic-based factor models. In the first stage they sort assets into portfolios based on book-to-price and market value characteristics. They use the differences between returns on the top and bottom fractile portfolios as proxies for the factor returns. They also include a market factor proxied by the return on a capitalization-weighted market index. In the second stage, the factor betas of portfolios and/or assets are estimated by time-series regression of asset returns on the derived factors. Carhart (1997) and Jegadeesh and Titman (1993, 2001) show that the addition of a momentum factor (proxied by high-twelve-month return minus low twelve-month-return) adds explanatory power to the Fama-French three-factor model, both in terms of explaining comovements and mean returns. Ang, Hodrick, Xing and Zhang (2006 a,b) and Goyal and Santa-Clara (2003) also find evidence for a own-volatility-related factor, both for explaining return comovements and mean returns.

One of the most empirically powerful factor decompositions for equity returns is an error-components model using industry affiliations. This involves setting the factor beta matrix equal to zero/one dummies, with row i containing a one in the j^{th} column if and only if firm i belongs to industry j. This is the simplest type of characteristic-based factor model of equity returns.

The first statistical factor is dominant in equity returns, accounting for 80-90% of the explanatory power in a multi-factor model. The standard specification of a error-components model does not isolate the "first" factor since its influence is spread across the factors. Heston and Rouwenhorst (1994) describe an alternative specification in which this factor is separated from the k industry factors. They add a constant to the model, so that the expanded set of factors $k+1$ is not directly identified (this lack of identification is sometimes called the "dummy variable trap," referring to a model that includes a full set of zero-one dummies plus a constant). Then, Heston and Rouwenhorst, impose an adding-up restriction on the estimated $k+1$ factors: the set of industry factors must sum to zero. This adding-up restriction on the factors restores statistical identification to the model, requiring constrained least squared in place of standard least squares estimation. It also provides a useful interpretation of the estimated factors: the factor associated with the constant term is the "market-wide" or "first" factor, and the factors associated with the industry dummies are the extra-market industry factors.

Heston and Rouwenhorst's adding-up condition is particularly useful in a multi-country context. It allows one to include an overlapping set of country and industry dummies without encountering the problem of the dummy variable trap. Including a constant, an international industry-country factor model must impose adding-up conditions both on the estimated industry factors and on the estimated country factors. This type of country-industry specification is useful for example in measuring the relative contribution of cross-border and national influences to return comovements, see, for example, Hopkins and Miller (2001).

6 Mutual Fund Separation

Given a set of traded assets and investors with different tastes for risk, it might be the case that each investor's portfolio is customized to the preferences of that investor and is quite different in composition than another investor's portfolio. When individual investors' optimal portfolio choices across all n assets can be reduced to the choice of combining $k < n$ portfolios, we say that k-fund separation holds. With k-fund separation, each investor is indifferent between choosing from the full set of n assets or from the set of k mutual funds. Tobin (1958) proves a one-fund separation theorem for investors choosing between cash and an array of money market instruments.

Separation results can be obtained either for particular specifications of agents' utility functions or for certain specifications of return distributions. Cass and Stiglitz (1970) derive necessary and sufficient conditions on utility functions to get separation. Given our emphasis of the role of factor models in portfolio selection and asset pricing, we will focus on the relation between return distributions and mutual fund separation. Ross (1978b) studies the conditions on return distributions that lead to mutual fund separation. Let x denote the payoff vector of a set of n assets next period. Weak k-fund separation holds if there exist k mutual funds, $\alpha^1, \alpha^2, ..., \alpha^k$, such that, for any portfolio, θ, and any monotonically increasing and concave (utility) function, u, there exists a portfolio α

$$\alpha = a_1\alpha^1 + a_2\alpha^2 + ... + a_k\alpha^k$$

with

$$E\{u(\alpha x)\} \geq E\{u(\theta x)\}.$$

Ross (1978b, Theorem 3) shows that necessary and sufficient conditions for weak k-fund separation are that returns follow a k-factor data generating process, that expected returns on assets are linearly dependent on their factor exposures, and that the k mutual funds have no non-factor risk. Thus, k-fund separation requires a restriction on the vector of mean returns in addition to restrictions on the covariance matrix of returns. The restriction on expected returns is inherently an asset pricing question, which we discuss in the next section.

7 Asset Pricing Theory

Mean variance analysis is a cornerstone of Asset Pricing Theory. Let x denote the payoff vector of a set of n assets next period, $p(x)$ denote the price of x and X denote the set of all payoffs that an investor can purchase. Assume that markets are competitive and frictionless, in the sense that there are no transactions costs, indivisibilities, or short sale constraints. The law of one price is the condition that two identical payoffs must have the same price. Thus, if payoff y is the same as a units of payoff x and b units of payoff z, then $p(y) = a \times p(x) + b \times p(z)$. Given the assumption of competitive and frictionless markets and the assumption of the law of one price (linearity of the pricing function) then there exists a unique payoff $x^* \in X$ for which the price of any payoff x in X is given by:

$$p(x) = E(x^*x) \tag{5}$$

(see Cochrane (2001, 4.1) for a proof). The random variable x^* is often referred to as a discount factor, pricing kernel, or state-price density. There may be other discount factors that are not in X which also price assets. In fact, any discount factor, m, with $m = x^* + \nu$ and $E(\nu x) = 0$ will also price all payoffs in X since $E(mx) = E(x^*x) + E(\nu x) = p(x) + 0 = p(x)$.

Define an arbitrage opportunity as the existence of a payoff, x, that is non-negative ($x \geq 0$) and is positive with positive probability (i.e., $\Pr(x > 0) > 0$), and which has a non-positive price ($p(x) \leq 0$). In a market where agents prefer more to less, we should not expect arbitrage opportunities to exist. The absence of arbitrage opportunities implies that the discount factor is strictly positive (see Ross (1978a) and Cochrane (2001)).

Given a valid traded discount factor, $m = x^*$, if a riskless asset exists with rate of return equal to r_f and gross payoff $R_f = 1 + r_f$, then $p(R_f) = 1 = E(mR_f) = E(m)R_f$. Therefore,

$$R_f = \frac{1}{E(m)}.$$

Since $E(mx) = E(m)E(x) + cov(m, x)$, we have that:

$$p(x) = \frac{E(x)}{R_f} + cov(m, x). \tag{6}$$

The price of payoff x is its expected value discounted at the riskless rate plus a risk premium or discount, $cov(m, x)$. Denote the gross (net) return on x as R_x (r_x). $R_x = \frac{x}{p(x)}$.and $r_x = \frac{x}{p(x)} - 1$. By dividing both sides of equation (6) by $p(x)$ and rearranging, we can express the pricing relation as:

$$E(r_x) - r_f = E(R_x) - R_f = -R_f \times cov(m, r_x). \tag{7}$$

From equations (6) and (7) we see that the only risk which influences asset prices (or equivalently expected returns) is the component of asset returns that is correlated with the discount factor, m. Asset expected returns can be expressed in terms of their "betas" relative to the discount factor by multiplying and dividing (7) by $var(m)$ and rearranging to get

$$E(r_x) - r_f = E(R_x) - R_f = \beta_{x,m}\lambda_m \tag{8}$$

where

$$\beta_{x,m} = \frac{cov(r_x, m)}{var(m)}$$
$$\lambda_m = \frac{-var(m)}{E(m)}$$

Rearranging (8) we find that

$$E(R_x) - R_f = \rho_{x,m} \frac{\sigma(m)}{E(m)} \sigma(R_x)$$

where $\rho_{x,m}$ is the correlation between asset returns and the discount factor. Since $\rho_{x,m}$ is between -1 and 1 we have the following bound on excess returns

$$|E(R_x) - R_f| \leq \frac{\sigma(m)}{E(m)} \sigma(R_x)$$

All assets must lie on, or to the right of, the two rays emanating from R_f as shown in the cross-hatched area in Figure 1. These rays are called the mean-variance boundary since they represent the lowest standard deviation possible for a given level of expected return. The upward-sloping portion of the mean-variance boundary is often call the efficient frontier since these portfolios have the highest expected return for a given level of standard deviation. Any portfolio on the mean-variance boundary is perfectly correlated with the discount factor. Let R^* denote a frontier portfolio, $R^* = x^*/p(x)$. We can write the discount factor as a linear function of R^*.

So far, we have invoked minimal assumptions to derive (8). In particular, we have not assumed specific forms of agents' utility functions, or properties of the distributions of asset returns other than the existence of second moments. We have assumed frictionless, competitive markets and the lack of arbitrage opportunities. While identifying portfolios on the mean-variance boundary, *ex post,* is relatively straight-forward, doing so *ex ante* is not. Much of the asset pricing literature is devoted to placing enough structure on the economic problem to derive the identity of m. We will begin by discussing consumption based models and then factor models of returns.

Assume that agents wish to maximize expected lifetime utility of the following sort

$$E_t \sum_{j=0}^{\infty} \delta^j u(c_{t+j}) \tag{9}$$

and that an asset is a claim to a stream of future payments, $d_{t+1}, d_{t+2}, \ldots$ The first order conditions of the maximization problem imply that the price of the asset is given by

$$p_t = E_t \sum_{j=1}^{\infty} \delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)} d_{t+j}$$

(see Lucas (1978), Brock (1982), and Cochrane (2001)). This can be re-written as

$$\begin{aligned} p_t \quad = \quad & E_t \left[\delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)} (d_{t+1} + p_{t+1}) \right] = \\ & E_t \left[m_{t+1} x_{t+1} \right]. \end{aligned} \tag{10}$$

Therefore, the consumption-based discount factor is $m_{t+1} = \delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)}$ which prices the payoff to holding the asset one period, $x_{t+1} = d_{t+1} + p_{t+1}$. Assets whose payoffs have higher covariance with future marginal utility (i.e., those that pay well in "bad times" when consumption is low and marginal utility is high) have higher values or, equivalently, lower expected returns. Dividing both sides of (10) by p_t yields the moment condition

$$1 = E_t\left[\delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)} \frac{(d_{t+1} + p_{t+1})}{p_t}\right] = E_t\left[\delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)} R_{x,t+1}\right] \tag{11}$$

which implies

$$E_t\left[\delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)} R_{x,t+1} - 1\right] = 0. \tag{12}$$

The moment condition in (12) can be multiplied by any instrument in the time t information set, say z_t, to get additional moment restrictions

$$E_t\left[\left(\delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)} R_{x,t+1} - 1\right) \times z_t\right] = 0. \tag{13}$$

Hansen and Singleton (1982) provide one of the earliest empirical tests of the restrictions in equation (13) assuming constant relative risk aversion, $u(c) = \frac{c^\gamma}{\gamma}$ for $\gamma < 1$. When equity and bond returns are included the restrictions implied by the pricing model are rejected. The low correlation between changes in consumption and stock returns requires very high levels of risk aversion to match the return on stocks. With expected utility, the high level of risk aversion implies a low level of intertemporal substitution which implies interest rates much higher that we observe in the data (see Hansen and Singleton (1982), Mehra and Prescott (1985), Kocherlakota (1996), Cochrane (2005)). The failure of the early tests of consumption-based asset pricing models might be due to properties of the consumption data, such as temporal aggregation and measurement error Breeden, Gibbons, and Litzenberger (1989). Alternatively, the specifications of the utility function may be inappropriate. Nonseparable utility specifications, such as habit formation, where utility is defined relative to a reference level of utility based on past consumption, seem to have some success in explaining average returns. Epstein and Zin (1989, 1991) derive and test a model in which utility is not state-separable. Their utility specification separates the level of risk aversion from the elasticity of substitution, and hence can fit the empirical regularities of asset returns better than the time- and state-separable specifications. Incorporating shocks to labor or entrepreneurial income, whether tradable (see Jagannathan and Wang (1996) and Campbell (1996)) or non-tradable (see Heaton and Lucas (1996, 1999) and Brav, Constantinides , and Geczy (2002)) help in matching mean returns.

Consumption-based asset pricing provides strong economic intuition about the determinants of risk premia. However, the challenges of working with consumption data are formidable. An alternative is suggested by the pricing function, (5), and the linear "beta" relation, (8). The pricing function shows that there is a portfolio whose payoff can serve as the discount factor. This portfolio is on the mean-variance boundary in Figure 1. In fact, any portfolio on the boundary will work. In terms of the consumption-based discount factor, say $m_{t+1} = \delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)}$, the portfolio of assets whose payoffs are maximally correlated with $\delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)}$ will price all assets. Thus $x^*$is the projection of $\delta^j \frac{u\prime(c_{t+j})}{u\prime(c_t)}$ onto the space of payoffs, X.

The earliest asset pricing literature began by making sufficient assumptions on preferences and/or the distribution of payoffs to identify x^*. With quadratic preferences or joint multivariate normality of returns, myopic preferences, and the lack of other sources of income besides the return on agents' portfolios, we obtain one-fund separation. Given that all agents hold the same mutual fund of risky assets, equilibrium between supply and demand for assets requires that that portfolio be the value-weighted market portfolio. This yields the Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1966), Treynor (1961, 1999), and Mossin (1966). Let r_M denote the rate of return on the market portfolio. The CAPM implies that $x^* = a + br_M$ and that

$$E(r_x) - r_f = E(R_x) - R_f = \beta_{x,M}\lambda_M \tag{14}$$

where

$$\begin{aligned} \beta_{x,M} &= \frac{cov(r_x, r_M)}{var(r_M)} \\ \lambda_M &= E(r_M) - r_f. \end{aligned}$$

The Arbitrage Pricing Theory (APT) of Ross (1976) assumes that asset returns follow a factor model, as in equation (1). If the factor model is a noiseless factor model (i.e., $\varepsilon \equiv 0$ in (1)), then the law of one price implies that expected returns are linear in assets exposure to the underlying factors. Equivalently the discount factor is a linear function of the factors:

$$\begin{aligned} x^* &= a + b'f \\ E(r_x) - r_f &= \beta'_{x,f}\lambda_f \end{aligned} \tag{15}$$

where $\beta_{x,f} = cov(r_x, f)$, a $k \times 1$ vector of covariances of r_x with the k factors, and $\lambda_f = -R_f cov(ff')b = -R_f E(x^* f)$.[3]

If the factor model is not noiseless, in the sense that ε is a mean-zero random vector, then we need sufficient conditions for the k mutual funds $\alpha^1, \alpha^2, ..., \alpha^k$, discussed in the section on mutual fund separation, to be well-diversified:

$$\alpha^j \varepsilon = 0$$

[3]See Cochrane (2001, section 6.3).

for $j = 1, 2, ..., k$. Connor (1984) derives an equilibrium version of the APT in which there are an infinite number of assets. In this economy we get $k-$fund separation and multi-factor pricing, as in (15) if the market portfolio is well-diversified.

With a strict factor model and a finite set of assets, we would, in general, not be able to construct k well-diversified mutual funds, so the APT pricing relation, equation (15) will hold as an approximation as derived in Ross (1976).

Alternative multi-factor asset pricing models include the Intertemporal Capital Asset Pricing Model (ICAPM) of Merton (1973) in which market risk and the risks associated with shifts in the consumption/investment opportunity set command risk premia. Empirical implementation of macroeconomic factor models often are built by postulating a set of variables that are good proxies for such risks. Example are Chan, Chen, and Hsieh (1985), Chen, Roll, and Ross (1986), Shanken (1990), Fama and French (1993, 1996), Carhart (1997). More recently there has been increased interest in estimating models that include factors that proxy for shocks to liquidity (e.g., Pástor and Stambaugh (2003), Acharya and Pedersen (2005), Korajczyk and Sadka (2008)). However, these models step away from the frictionless markets assumption underlying the derivation of the discount factor.

These empirical implementations of factor models essentially take a stand on the set of macroeconomic variables that drive a traded discount factor x^* or a potentially non-traded discount factor $m = x^* + \nu$.

8 Summary

The mean-variance analysis of Harry Markowitz provides the basis for portfolio and asset pricing theory. Within those large literatures, factor models have played a significant role. In portfolio theory, strict factor models impose structure on covariance matrices in ways that make portfolio optimization feasible in situations where unconstrained problems are infeasible or computationally burdensome.

The role of mean-variance boundary portfolios is central in asset pricing. With competitive and frictionless markets and the existence of second moments, the law of one price implies the existence of a traded portfolio whose payoff can be used to price all assets. The portfolio is on the mean-variance boundary (the set of portfolios with minimum variance for each level of expected return). This portfolio payoff will price all assets, including derivative securities, regardless of the distribution of asset returns or the preferences of investors (again assuming preferences rule out arbitrage opportunities and the relevant moments exist).

Empirical applications of asset pricing theory, of course, require the identification of the appropriate discount factor. While consumption-based models provide a great deal of economic intuition, the extant consumption data often are poor proxies for the quantities needed to apply the models. Factor models, whether statistical, macroeconomic, or characteristic-based, provide different paths to creating sets of portfolios that provide candidate discount factors.

References

[1] Acharya, Viral V., and Lasse Heje Pedersen, 2005, "Asset Pricing with Liquidity Risk," *Journal of Financial Economics* 77, 375-410.

[2] Ang, Andrew, Robert J. Hodrick, Yuhang Xing and Xiaoyan Zhang, 2006a, "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61, 259-299.

[3] Ang, Andrew, Robert J. Hodrick, Yuhang Xing and Xiaoyan Zhang, 2006b, "High Idiosyncratic Risk and Low Returns: International and Further U. S. Evidence," Working paper, Columbia Business School, Columbia University.

[4] Bai, Jushan, 2003, "Inferential Theory for Factor Models of Large Dimension," *Econometrica* 71, 135-171.

[5] Bai, Jushan, and Serena Ng, 2002, "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70, 191-221.

[6] Basilevsky, Alexander, 1994, *Statistical Factor Analysis and Related Methods: Theory and Applications*. New York: Wiley.

[7] Brav, Alon, George M. Constantinides, Christopher C. Geczy, 2002, "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence," *Journal of Political Economy* 110, 793-824.

[8] Breeden, Douglas T., Michael R. Gibbons, Robert H. Litzenberger, 1989, "Empirical Test of the Consumption-Oriented CAPM," *Journal of Finance* 44, 231-262

[9] Brock, William A., 1982, "Asset Prices in a Production Economy," in *The Economics of Information and Uncertainty* edited by John J. McCall. Chicago: University of Chicago Press.

[10] Campbell, John Y., 1996, "Understanding Risk and Return" *Journal of Political Economy* 104, 298-345.

[11] Campbell, John Y., 2003, "Consumption-Based Asset Pricing" in *Handbook of the Economics of Finance*, edited by G M Constantinides, M Harris and R Stulz. Amsterdam: North Holland.

[12] Carhart, Mark M., 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57-82.

[13] Cass, David, and Joseph E. Stiglitz, 1970, "The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds," *Journal of Economic Theory* 2, 122-160.

[14] Chamberlain, Gary, and Michael Rothschild, 1983, "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica* 51, 1305-1324.

[15] Chan, K. C., Nai-fu Chen, and David Hsieh, 1985, "An Exploratory Investigation of the Firm Size Effect." *Journal of Financial Economics* 14, 451-471.

[16] Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business* 59, 383-403.

[17] Cochrane, John H., 2001, *Asset Pricing*, Princeton: Princeton University Press.

[18] Cochrane, John H., 2005, "Financial Markets and the Real Economy," *Foundations and Trends in Finance* 1, 1-101.

[19] Cohen, Kalman J. and Jerry A. Pogue, 1967, "An Empirical Evaluation of Alternative Portfolio-Selection Models," *Journal of Business* 40, 166-193.

[20] Connor, Gregory, 1984, "A Unified Beta Pricing Theory," *Journal of Economic Theory* 34, 13-31.

[21] Connor, Gregory, 1995, "The Three Types of Factor Models: A Comparison of Their Explanatory Power," *Financial Analysts Journal* 51, 42-46.

[22] Connor, Gregory, and Korajczyk, Robert A., 1986, "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of Financial Economics* 15, 373-394.

[23] Connor, Gregory, and Korajczyk, Robert A., 1991, "The Attributes, Behavior, and Performance of U.S. Mutual Funds." *Review of Quantitative Finance and Accounting* 1, 5-26.

[24] Connor, Gregory, and Korajczyk, Robert A., 1993, "A Test for the Number of Factors in an Approximate Factor Model." *Journal of Finance* 48, 1263-1291.

[25] Connor, Gregory, and Korajczyk, Robert A., 2007, "Factor Models of Asset Returns," Forthcoming, *Encyclopedia of Quantitative Finance*, edited by Rama Cont. Chicester: Wiley, (http://ssrn.com/abstract=1024709).

[26] Daniel, Kent, and Sheridan Titman, 1997, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance* 52, 1-33.

[27] Dhrymes, Phoebus J., Irwin Friend and N. Bulent Gultekin, 1984, "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory, *Journal of Finance* 39, 323-246.

[28] Elton, Edwin J. and Martin J. Gruber, 1973, "Estimating the Dependence Structure of Share Prices - Implications for Portfolio Selection," *Journal of Finance* 28, 1203-1232.

[29] Elton, Edwin J., Martin J. Gruber, and Manfred W. Padberg, 1979, "Simple Criteria for Optimal Portfolio Selection: The Multi-Index Case." In E. J. Elton and M. J. Gruber (eds.) *Portfolio Theory, 25 Years After: Essays in Honor of Harry Markowitz*. Amsterdam: North-Holland.

[30] Epstein, Larry G., and Stanley E. Zin, 1989, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica* 57, 937-969.

[31] Epstein, Larry G., and Stanley E. Zin, 1991, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy* 99, 263-286.

[32] Fama, Eugene F., and Kenneth R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47, 427-465.

[33] Fama, Eugene F., and Kenneth R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3-56.

[34] Fama, Eugene F., and Kenneth R. French, 1996, "Multifactor explanations of asset pricing anomalies," *Journal of Finance* 51, 55-84.

[35] Farrar, Donald E., 1962, *The investment decision under uncertainty*, Englewood Cliffs: Prentice-Hall.

[36] Goyal, Amit, and Pedro Santa-Clara, 2003, "Idiosyncratic risk matters!", *Journal of Finance* 58, 975-1007.

[37] Hansen, Lars Peter, and Kenneth J. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269-1286.

[38] Heaton, John, and Deborah J. Lucas, 1996, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy* 104, 443-487.

[39] Heaton, John, and Deborah J. Lucas, 1999, "Stock Prices and Fundamentals," *NBER Macroeconomics Annual* 14 213-242.

[40] Heston, Steven L., and K. Geert Rouwenhorst, 1994, "Does Industrial Structure Explain the Benefits of International Diversification?" *Journal of Financial Economics* 36, 3-27.

[41] Hopkins, Peter J. B. and C. Hayes Miller, 2001, *Country, Sector and Company Factors in Global Equity Models.* The Research Foundation of AIMR and the Blackwell Series in Finance, Charlottesvile VA.

[42] Huang, Chi-fu and Robert H., Litzenberger, 1988, *Foundations for Financial Economics,* New York: Elsevier Science.

[43] Jagannathan, Ravi, and Zhenyu Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance* 51, 3-53.

[44] Jegadeesh, Narasimhan, and Sheridan Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance* 48, 65-91.

[45] Jegadeesh, Narasimhan, and Sheridan Titman, 2001, "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations," *Journal of Finance* 56, 699-718.

[46] King, Benjamin F., 1966, "Market and Industry Factors in Stock Price Behavior," *Journal of Business* 39, 139-190.

[47] Kocherlakota, Narayana R., 1996, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature* 34, 42-71.

[48] Korajczyk, Robert A. and Ronnie Sadka, 2008, "Pricing the Commonality Across Alternative Measures of Liquidity," *Journal of Financial Economics* 87: 45-72.

[49] Lintner, John, 1965, "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance* 20, 587-615.

[50] Lucas, Robert E., Jr., 1978, "Asset Prices in an Exchange Economy," *Econometrica* 46, 1429-1446.

[51] Markowitz, Harry M., 1952, "Portfolio Selection." *Journal of Finance* 7, 77-91.

[52] Markowitz, Harry M., 1959, *Portfolio Selection: Efficient Diversification of Investments*, New York: Wiley.

[53] Markowitz, Harry M., 1984, "The Two-Beta Trap." *Journal of Portfolio Management* 11, 12-19.

[54] Markowitz, Harry M., 1987, *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Oxford: Basil Blackwell.

[55] Mehra, Rajnish, and Edward C. Prescott, 1985, " The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15, 145-161.

[56] Merton, Robert C., 1973. "An Intertermporal Capital Asset Pricing Model," *Econometrica* 41, 867-887.

[57] Mossin, Jan, 1966, "Equilibrium in a Capital Asset Market," *Econometrica* 34, 768-783.

[58] Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory.* New York: Wiley.

[59] Pástor, Ľuboš, and Robert F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy 111*, 642-685.

[60] Roll, Richard and Ross, Stephen A., 1980, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance* 35, 1073-1103.

[61] Rosenberg, Barr, 1974, "Extra-Market Components of Covariance in Security Returns," *Journal of Financial and Quantitative Analysis* 9, 263-274.

[62] Ross, Stephen A., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* 13, 341-360.

[63] Ross, Stephen A., 1978a, "A Simple Approach to the Valuation of Risky Streams," *Journal of Business* 51, 453-475.

[64] Ross, Stephen A., 1978b, "Mutual Fund Separation in Financial Theory — The Separating Distributions," *Journal of Economic Theory* 17, 254-286.

[65] Shanken, Jay, 1987, "Nonsynchronous data and the covariance-factor structure of returns," *Journal of Finance* 42, 221-231.

[66] Shanken, Jay, 1990, "Intertemporal Asset Pricing: An Empirical Investigation," *Journal of Econometrics* 45, 99-120.

[67] Shanken, Jay, and Mark I. Weinstein, 2006, "Economic Forces and the Stock Market Revisited," *Journal of Empirical Finance* 13, 129-144.

[68] Sharpe, William F., 1963, A simplified model for portfolio analysis, *Management Science* 9, 277-293.

[69] Sharpe, William F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19, 425-442.

[70] Stock, James H. and Watson, Mark W., 2002, "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics* 20, 147-162.

[71] Stroyny, Alvin L., 2005, "Estimating a Combined Linear Factor Model." In *Linear Factor Models in Finance,* edited by John Knight and Stephen Satchell. Oxford: Elevier Butterworth-Heinemann.

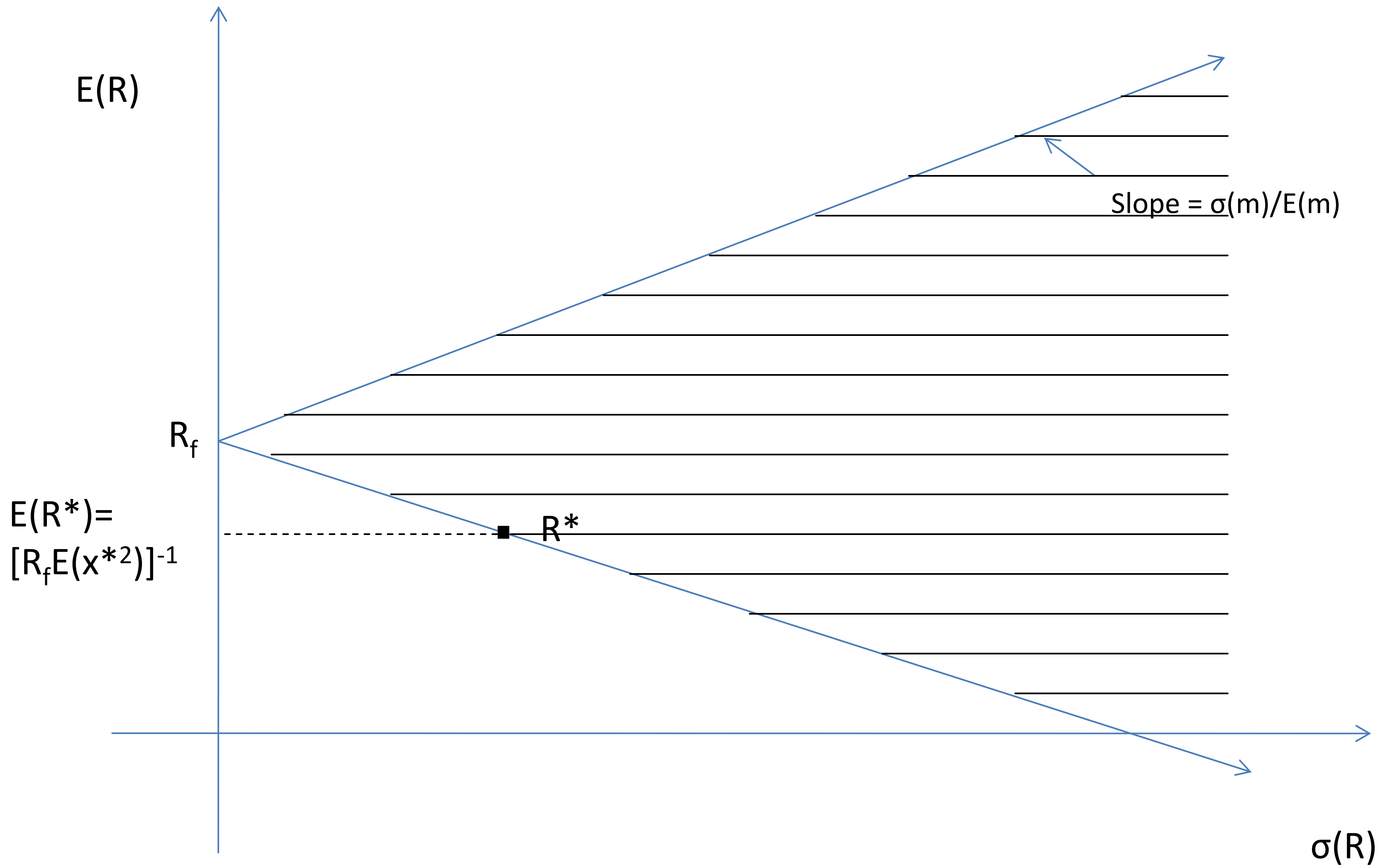
[72] Tobin, J., 1958, "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies* 25, 65-86.

[73] Treynor, Jack L., 1961, "Toward a Theory of Market Value of Risky Assets." Unpublished Working Paper.

[74] Treynor, Jack L., 1999, "Towards a Theory of Market Value of Risky Assets." In *Asset Pricing and Portfolio Performance: Models, Strategy, and Performance Metrics,* edited by Robert A. Korajczyk. London: Risk Publications.

[75] Vassalou, Maria, 2003, "News Related to Future GDP Growth as a Risk Factor in Equity Returns," *Journal of Financial Economics* 68, 47–73.

[76] Weil, Philippe, 1989, "The Equity Premium Puzzle and the Risk-free Rate Puzzle," *Journal of Monetary Economics* 24, 401-421.



E(R)
Slope = σ(m)/E(m)
R_f
E(R*)=
$[R_f E(x^{*2})]^{-1}$
R*
σ(R)

Figure 1: The cross-hatched area represents the set of attainable expected return/Standard deviation combinations. The mean-variance boundary can be obtained by combinations of R^* and R_f.