NBER WORKING PAPER SERIES

MULTIFACTOR MODELS DO NOT EXPLAIN DEVIATIONS FROM THE CAPM

A. Craig MacKinlay

Working Paper No. 4756

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1994

I thank Fischer Black, Kent Daniel, Eugene Fama, Bruce Grundy, John Heaton, Ravi Jagannathan, Shmuel Kandel, Krishna Ramaswamy, Bill Schwert, Jay Shanken, Robert Whitelaw, and especially Andrew Lo and Robert Stambaugh for useful discussions relating to this paper. A preliminary draft was completed while I was a summer visitor at the University of British Columbia in July 1991. Thanks to the UBC finance faculty for constructive suggestions. An early draft circulated under the title "Distinguishing Among Asset Pricing Theories." I am grateful to seminar participants at Indiana University, MIT, NBER Asset Pricing Conference, Princeton, Rutgers University, Stanford, University of California at Berkeley, University of Chicago, University of Illinois, University of Rochester, Vanderbilt, and Wharton for helpful comments. Research support from the Batterymarch Fellowship and the Geewax-Terker Research Fund is gratefully acknowledged. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #4756
June 1994

MULTIFACTOR MODELS DO NOT EXPLAIN DEVIATIONS FROM THE CAPM

ABSTRACT

A number of studies have presented evidence rejecting the validity of the Capital Asset Pricing Model (CAPM). This evidence has spawned research into possible explanations. These explanations can be divided into two main categories - the risk based alternatives and the nonrisk based alternatives. The risk based category includes multifactor asset pricing models developed under the assumptions of investor rationality and perfect capital markets. The nonrisk based category includes biases introduced in the empirical methodology, the existence of market frictions, or explanations arising from the presence of irrational investors. The distinction between the two categories is important for asset pricing applications such as estimation of the cost of capital. This paper proposes to distinguish between the two categories using *ex ante* analysis. A framework is developed showing that *ex ante* one should expect that CAPM deviations due to missing risk factors will be very difficult to statistically detect. In contrast, deviations resulting from nonrisk based sources will be easy to detect. Examination of empirical results leads to the conclusion that the risk based alternatives is not the whole story for the CAPM deviations. The implication of this conclusion is that the adoption of empirically developed multifactor asset pricing models may be premature.

A. Craig MacKinlay
Department of Finance
Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6367
and NBER

1 Introduction

One of the important problems of modern financial economics is the quantification of the tradeoff between risk and expected return. Although common sense suggests that investments free of risk will generally yield lower returns than more risky investments such as the stock market, it was only with the development of the Sharpe-Lintner Capital Asset Pricing Model (CAPM) that economists were able to say how much lower this would have to be. In particular, the CAPM shows that the cross-section of expected excess returns of financial assets must be linearly related to the market betas with an intercept of zero. Because of the practical importance of this risk/return relation, numerous studies have empirically examined this implication. Over the past fifteen years, a number of these studies have presented evidence which contradict the CAPM in that the hypothesis that the intercept of a regression of excess returns on the excess return of the market is zero is statistically rejected.

The apparent violations of the CAPM have spawned research into possible explanations.[1] For the analysis of this paper the explanations will be divided into two categories – the risk based alternatives and the nonrisk based alternatives. The risk based category includes multifactor asset pricing models developed under the assumptions of investor rationality and perfect capital markets. For this category the source of deviations from the CAPM is missing risk factors. The nonrisk based category includes biases introduced in the empirical methodology, the existence of market frictions, or explanations arising from the presence of irrational investors.[2] Examples are data–snooping biases, biases in computing returns, transaction costs and liquidity effects, and market inefficiencies.

The finding of empirical tests of the CAPM that the intercepts deviate statistically from zero has naturally lead to the empirical examination of multifactor asset pricing models motivated by the Arbitrage Pricing Theory (APT) developed by Ross (1976) and the Intertemporal Capital Asset Pricing Model (ICAPM) developed by Merton (1973).[3] The basic

[1] Of course, Roll's (1977) critique is relevant here. In a strict sense, most studies reject the mean–variance efficiency of the CRSP Indexes. In the analysis of this paper, the missing risk factors analysis will apply to the case where the source of the violations is the misidentification of the market portfolio.

[2] Although this category is labelled nonrisk based, clearly some of the explanations contain elements of risk.

[3] See Fama (1993) for detailed discussion of these multifactor model theories.

approach has been to introduce additional factors in the form of excess returns on traded portfolios and then re-examine the zero-intercept hypothesis. A recent example of a paper which includes this approach is Fama and French (1993). They document that the estimates of the CAPM intercepts deviate from zero for portfolios formed on the basis of book value to market value of equity ratios as well as for portfolios formed based on market capitalization.[4] Upon finding that the intercepts for these portfolios with a three factor model are closer to zero, they conclude that missing risk factors in the CAPM are the source of the deviations. They go on to advocate the use of a multifactor model, stating that, with respect to the use of the Sharpe–Lintner CAPM, their results "should help to break this common habit" (Fama and French (1993) page 44).

However, the conclusion that additional risk factors are required may be premature. There are a number of other explanations which are consistent with the presence of deviations. One explanation attributes the CAPM deviations to data–snooping. This explanation is presented in Lo and MacKinlay (1990). The argument is that on an *ex post* basis one will always be able to find deviations from the CAPM. Such deviations considered in a group will appear statistically significant. However, under this explanation they are not real but just a result of grouping assets with common disturbance terms. Since, in financial economics our empirical analysis is *ex post* in nature, this problem is difficult to directly control. Further, in practice, direct adjustments for potential snooping are difficult to implement and, when implemented, make it very difficult to find real deviations.

While, in general, it is difficult to quantify and adjust for the effects of data–snooping biases, there are some related biases which can be examined. One such case pursued by Kothari, Shanken, and Sloan (1993) is sample selection bias. These authors show that significant biases can arise in academic research when the analysis is conditioned on the assets appearing in both the CRSP database and the COMPUSTAT database. Their analysis suggests that deviations from the CAPM such as those documented by Fama and French (1993) can be explained by sample selection biases. Breen and Korajczyk (1993) provide

[4]Fama and French are also cocerned with the observation that the relation between average returns and market betas is weak. This point is not addressed in this paper but has been addressed in a number of recent papers. Some examples are Chan and Lakonishok (1992), Kandel and Stambaugh (1993), Kothari, Shanken and Sloan (1993), and Roll and Ross (1992).

further evidence on selection biases providing support for the Kothari, Shanken, and Sloan conclusion.

Other researchers interpret the deviations from the CAPM as indications of the presence of irrational behavior by market participants. A number of theories have been developed which are consistent with this line of thought.[5] A recent example is the work of Lakonishok, Shleifer, and Vishny (1993) where they argue that the deviations arise from investors following naive strategies including extrapolating past growth rates too far into the future, assuming a trend in stock prices, overreacting to good or bad news, or liking to invest in firms with a high level of profitability. With this alternative the possibility of non-zero intercepts arises not only from missing risk factors but from specific firm characteristics.

Conrad and Kaul (1993) consider the possibility that biases in computed returns explain the deviations. They find that the implicit portfolio rebalancing in most analyses biases measured returns upwards. This leads to overstated returns and consequently CAPM deviations. This problem will be the most severe for tests using frequently rebalanced portfolios and short observation intervals.

Finally, market frictions and liquidity effects could induce a non-zero intercept in the CAPM tests. Since the model is developed in a perfect market, such effects are not accommodated. Amihud and Mendleson (1986) present some evidence of returns containing effects from market frictions and demands for liquidity. Also, Luttmer (1991) provides evidence on the importance of market frictions in the context of the bounds on the moments of the intertemporal marginal rate of substitution derived in Hansen and Jagannathan (1991).

The controversy of whether or not the CAPM deviations are due to missing risk factors flourishes because empirically it is hard to distinguish between the various arguments. The difficulty arises because, on an *ex post* basis, one can always find a set of risk factors that will make the asset pricing model intercept zero.[6] Given this, without specific theory identifying the risk factors, one will always be able to explain the cross-section of expected returns with a multifactor asset pricing model. This will be true even if the real explanation lies in one of the nonrisk based categories.

[5]DeBondt and Thaler (1985) is an important early paper responsible for initiating interest in this area.

[6]Roll (1977) argues this point for the CAPM case.

Depsite the difficulty in distinguishing between the risk based and nonrisk based categories, the practical implications of the distinction are important. One application illustrating the importance is the estimation of the cost of capital to evaluate investment opportunites. If the risk based explanation is correct, then cost of capital calculations using the CAPM can be badly misspecified. A better approach would be to use a multifactor model which captures the missing risk factors. On the other hand, if the deviations are a result of the nonrisk based explanations then disposing of the CAPM in favor of a multifactor model may lead to serious errors. The cost of capital estimate from a multifactor model can be very different than the estimate from the CAPM.

In this paper I discriminate between the risk based and the nonrisk based explanations using *ex ante* analysis. The objective is to evaluate the plausability of the argument that the deviations from the CAPM can be explained by additional risk factors. I argue that *ex ante* one should expect that CAPM deviations due to missing risk factors will be very difficult to statistically detect. Intuitively this is because the deviation in expected return is also accompanied by increased variance. I formally analyze the issue using mean-variance efficient set mathematics in conjunction with the zero-intercept F-test presented in Gibbons, Ross, and Shanken (1989) and MacKinlay (1987). Technically the difficulty exists because when deviations from the CAPM or other multifactor pricing models are the result of omitted risk factors, there is an upper limit on the distance between the null distribution of the test statistic and the alternative distribution. With the nonrisk based alternatives where the source of the deviations is not missing factors, no such limit exists.

The paper also draws on a related distinction between the two categories. This distinction is the difference in the behavior of the maximum Sharpe measure squared as the cross-section of securities is increased.[7] For the risk based alternatives the maximum Sharpe measure squared is bounded and for the nonrisk based alternatives the maximum Sharpe measure squared is a less useful construct and can, in principle, be unbounded.

The results of the paper underscore the important role that economic analysis plays in distinguishing among different pricing models for the relation between risk and return. In

[7]The Sharpe measure is the ratio of the mean excess return to the standard deviation of the excess return.

the absence of specific alternative theories, without very long time series of data, one is limited in what can be said about risk/return relations among financial securities.

The paper proceeds as follows. In section 2 the framework for the analysis is presented. In section 3 the *optimal orthogonal portfolio* is defined. This portfolio will play a key role in the arguments of the paper. Many of the results in the paper can be related to the values of the squared Sharpe measure for relevant portfolios. In section 4 the relations between the parameters of the returns and the Sharpe measures are presented. Section 5 develops the implications relating to the missing risk factors controversy. Theoretically, the framework used to distinguish between the risk based and nonrisk based explanations assumes a large number of assets. Section 6 illustrates that the usefulness of the framework does not depend on this assumption. The paper concludes with section 7.

2 Linear Pricing Models and Mean-Variance Analysis.

We begin by specifying the distributional properties of excess returns for $\overline{N}$ primary assets in the economy. Let z_t represent the $\overline{N}$x1 vector of excess returns for period t. Assume z_t is stationary and ergodic with mean μ and covariance matrix V which is full rank. Given these assumptions for any set of factor portfolios, a linear relation between the excess returns and the portfolios' excess returns results. The relation can be expressed as:

$$z_t = \alpha + Bz_{pt} + \epsilon_t \tag{1}$$

$$E[\epsilon_t] = 0 \tag{2}$$

$$E[\epsilon_t \epsilon_t'] = \Sigma \tag{3}$$

$$E[z_{pt}] = \mu_p \quad , \quad E[(z_{pt} - \mu_p)(z_{pt} - \mu_p)'] = \Omega_p \tag{4}$$

$$Cov[z_{pt}, \epsilon_t] = 0. \tag{5}$$

B is the $\overline{N} \times K$-matrix of factor loadings, z_{pt} is the $K \times 1$-vector of time-t factor portfolio excess returns, and α and ϵ_t are $\overline{N} \times 1$-vectors of asset return intercepts and disturbances respectively.[8]

It is well-known that all of the elements of the vector α will be zero if a linear combination of the factor portfolios form the tangency portfolio (i.e. the mean–variance efficient portfolio of risky assets given the presence of a riskfree asset). Let z_{qt} be the excess return of the (ex ante) tangency portfolio and let x_q be the $\overline{N} \times 1$ vector of portfolio weights. Here, and throughout the paper, let ι represent a conforming vector of ones. From mean–variance analysis:

$$x_q = (\iota' V^{-1} \mu)^{-1} V^{-1} \mu. \tag{6}$$

In the context of our previous discussion, the asset pricing model will be considered well-specified when the tangency portfolio can be formed from a linear combination of the K-factor portfolios.

In the next section a portfolio is constructed which will be useful to characterize the asset pricing model deviations when factor portfolios are not jointly mean–variance efficient.

3 Optimal Orthogonal Portfolio.

Our interest is in formally developing the relation between the deviations from the asset pricing model, α, and the residual covariance matrix Σ when a linear combination of the factor portfolios do not form the tangency portfolio. To facilitate this I define the *optimal orthogonal portfolio*.[9] This is the unique portfolio which can be combined with the factor portfolios to form the tangency portfolio and is orthogonal to the factor portfolios. I next

[8]The dependence of α, B, and Σ on the factor portfolios is suppressed for notational convenience.

[9]See Roll (1980) for properties of orthogonal porfolios in a general context and see Lehmann (1987, 1988, 1992) for discussions of the role of orthogonal portfolios in asset pricing tests. Also related is the orthogonal factor employed in MacKinlay (1987), the active portfolio considered by Gibbons, Ross and Shanken (1989) and the modifying payoff used in Hansen and Jagannathan (1994).

formally define this portfolio.

Definition: *optimal orthogonal portfolio.*

Take as given K factor portfolios which cannot be combined to form the tangency portfolio or the global minimum variance portfolio. A portfolio h will be defined as the optimal orthogonal portfolio with respect to these K factor portfolios if:

$$x_q \;=\; X_p\,\omega \;+\; x_h\,(1-\iota'\omega) \tag{7}$$

and

$$x_h'\,V\,X_p \;=\; 0 \tag{8}$$

for a $(K \times 1)$ vector ω where X_p is the $(\overline{N} \times K)$ matrix of asset weights for the factor portfolios, x_h is the $(\overline{N} \times 1)$ vector of asset weights for the optimal orthogonal portfolio, and x_q is the $(\overline{N} \times 1)$ vector of asset weights for the tangency portfolio. If one considers a model without any factor portfolios ($K = 0$) then the optimal orthogonal portfolio will be the tangency portfolio.

The weights of portfolio h can be expressed in terms of the parameters of the one factor model. For the vector of weights:

$$\begin{aligned} x_h &= (\iota' V^{-1}\alpha)^{-1} V^{-1}\alpha \\ &= (\iota'\Sigma^{\dagger}\,\alpha)^{-1}\Sigma^{\dagger}\,\alpha. \end{aligned} \tag{9}$$

where the $\dagger$ superscript indicates the generalized inverse. The usefulness of this portfolio comes from the fact that when added to (1) the intercept will vanish and the factor loading matrix B will not be altered. The optimality restriction in (7) leads to the intercept vanishing, and the orthogonality condition in (8) leads to B being unchanged. Adding in z_{ht}:

$$z_t \;=\; Bz_{pt} \;+\; \beta_h z_{ht} \;+\; u_t \tag{10}$$

$$E[u_t] = 0 \tag{11}$$

$$E[u_t u_t'] = \Phi \tag{12}$$

$$E[z_{ht}] = \mu_h \quad , \quad E[(z_{ht} - \mu_h)^2] = \sigma_h^2 \tag{13}$$

$$Cov[z_{pt}, u_t] = 0. \tag{14}$$

$$Cov[z_{ht}, u_t] = 0. \tag{15}$$

The link results from comparing (1) and (10). Taking the unconditional expectations of both sides:

$$\alpha = \beta_h \mu_h \tag{16}$$

and by equating the variance of ϵ_t with the variance of $\beta_h z_{ht} + u_t$:

$$\begin{aligned} \Sigma &= \beta_h \beta_h' \sigma_h^2 + \Phi \\ &= \alpha\, \alpha' \frac{\sigma_h^2}{\mu_h^2} + \Phi. \end{aligned} \tag{17}$$

The key link between the model deviations and the residual variances and covariances emerges from (17). The intuition for the link is straightforward. Deviations from the model must be accompanied by a common component in the residual variance in order to prevent the formation of a portfolio with a positive deviation and a residual variance which decreases to 0 as the number of securities in the portfolio grows. In cases where the link is not present (i.e. the link is undone by Φ), asymptotic arbitrage opportunities will exist.

4 Squared Sharpe Measures.

The squared Sharpe measure is a useful construct for interpreting much of the ensuing analysis. The Sharpe measure for a given portfolio is calculated by dividing the mean excess return by the standard deviation of return. It is well-known that the tangency portfolio q will have the maximum squared Sharpe measure of all portfolios.[10] The squared Sharpe measure of q, s_q^2, is:

$$s_q^2 \;=\; \mu' V^{-1} \mu. \tag{18}$$

Since the K factor portfolios p and the optimal orthogonal portfolio h can be combined to form the tangency portfolio, it follows that the maximum squared Sharpe measure of these $K+1$ portfolios will be s_q^2. Since h is orthogonal to the portfolios p, one can express s_q^2 as the sum of the squared Sharpe measure of the orthogonal portfolio and the squared maximum Sharpe measure of the factor portfolios,

$$s_q^2 \;=\; s_h^2 \;+\; s_p^2 \tag{19}$$

where $s_h^2 \;=\; \frac{\mu_h^2}{\sigma_h^2}$ and $s_p^2 \;=\; \mu_p' \Omega_p^{-1} \mu_p$.

In applications I will be employing subsets of the $\overline{N}$ assets. Results similar to those above will hold within a subset of N assets. For the subset analysis when considering the tangency portfolio (of the subset), the maximum squared Sharpe measure of the assets and the optimal orthogonal portfolio for the subset, it is necessary to augment the N assets with the factor portfolios p. Defining $z_{t_s}^*$ as the $(N+K \times 1)$ vector $[z_t' \; z_{pt}']'$ with mean $\mu_s^{*\prime}$ and covariance matrix V_s^*, we have for the tangency portfolio of these $N+K$ assets:

$$s_{q_s}^2 \;=\; \mu_s^{*\prime} V_s^{*-1} \mu_s^*. \tag{20}$$

The subscript s indicates we are using a subset of the assets.

[10]See Jobson and Korkie (1982) for a development of this point and a performance measurement application. The existence of a maximum Sharpe measure as the number of asssets become large is central to the arbitrage pricing theory. For further discussion see Chamberlain and Rothschild (1983) and Ingersoll (1984).

As we shall see, the analysis (with a subset of assets) will involve the quadratic $\alpha'\Sigma^{-1}\alpha$ computed using the parameters for the N assets. Gibbons, Ross, and Shanken (1989) and Lehmann (1988, 1992) provide interpretations of this quadratic term in terms of Sharpe measures. Assuming Σ is of full rank,[11] they show:

$$\alpha_s'\Sigma_s^{-1}\alpha_s \;=\; s_{q_s}^2 \;-\; s_p^2. \tag{21}$$

Consistent with (19), for the subset of assets $\alpha'\Sigma^{-1}\alpha$ will be the squared Sharpe measure of the subset's optimal orthogonal portfolio h_s. Therefore for the a given subset of assets we have:

$$s_{h_s}^2 \;=\; \alpha_s'\Sigma_s^{-1}\alpha_s \tag{22}$$

and

$$s_{q_s}^2 \;=\; s_{h_s}^2 + s_p^2. \tag{23}$$

We also note that the squared Sharpe measure of the subset's optimal orthogonal portfolio is less than or equal to that of the population optimal orthogonal portfolio. We have:

$$s_{h_s}^2 \;\leq\; s_h^2. \tag{24}$$

Next we use the optimal orthogonal portfolio and the Sharpe measures results together with the model deviation residual variance link to develop implications for distinguishing among asset pricing models. Hereafter I will suppress the s subscript. No ambiguity will result, since, in the subsequent analysis, we will be working only with subsets of the assets.

[11] If Σ is singular then one must use the generalized inverse.

5 Implications for Risk Based Versus Nonrisk Based Alternatives.

Many asset pricing model tests involve testing the null hypothesis that the model intercept is zero using tests in the spirit of the zero-intercept F-test.[12] A common conclusion is that rejection of this hypothesis using one or more factor portfolios is an indication that more risk factors are required to explain the risk–return relation. This conclusion has lead to the inclusion of additional factors so that the null hypothesis will be accepted. A shortcoming of this approach is that, after adding factors, when all is said and done there are multiple potential interpretations of why the hypothesis is accepted. One view most recently advocated by Fama and French (1993) is that we have made genuine progress in terms of identifying the "right" asset pricing model. An alternative view is that, since the additional factors lack strong theoretical motivation, we have suceeded in finding a within sample fit through data–snooping. Certainly supporters of the nonrisk based positions would argue this alternative view.

In this section we employ *ex ante* analysis to attempt to discriminate between the two interpretations. The analysis integrates the link between the pricing model intercept and the residual covariance matrix of (17) and the squared Sharpe measures results with the distribution theory for the zero-intercept F-test. We consider two approaches. The first approach is a testing approach which compares the null hypothesis test statistic distribution with the distribution under each of the alternatives. The second approach is estimation based, drawing on the squared Sharpe measures analysis to develop estimators for the squared Sharpe measure of the optimal orthgonal portfolio. Before presenting the two approaches the zero-intercept F-test is summarized.

[12] Examples of tests which basically fit into this framework are those in Campbell (1987), Connor and Korajczyk (1988), Fama and French (1993), Gibbons, Ross, and Shanken (1989), Huberman, Kandel, and Stambaugh (1987), Lehmann and Modest (1988), and MacKinlay (1987). The arguments in the paper can also be related to the zero-beta CAPM tests in Gibbons (1982), Shanken (1985), and Stambaugh (1982).

5.1 Zero Intercept F-Test.

To implement the F-test the additional assumption that excess asset returns are jointly normal and temporally independently and identically distributed is added. This assumption, though restrictive, buys us exact finite sample distributional results thereby simplifying the analysis. However, it is important to note that this assumption is not central to the point; similar results will hold under much weaker assumptions. Using a Generalized Method of Moments approach, MacKinlay and Richardson (1991) present a more general test statistic which has asymptotically a chi-square distribution. Analysis similar to that presented for the F-test holds for this general statistic.

We begin with a summary of the zero-intercept F-test of the null hypothesis that the intercept vector α from (1) is be 0. Let H_o be the null hypothesis and H_a be the alternative.

$$H_o: \qquad \alpha = 0 \tag{25}$$

$$H_a: \qquad \alpha \neq 0. \tag{26}$$

H_o can be tested using the following test statistic:

$$\theta_1 = [(T-N-K)/N][1+\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p]^{-1}\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}. \tag{27}$$

where T is the number of time series observations, N is the number of assets or portfolios of assets included, and K is the number of factor portfolios. The hat superscripts indicate the maximum likelihood estimators. Under the null hypothesis, θ_1 is unconditionally distributed central F with N degrees of freedom in the numerator and $(T-N-K)$ degrees of freedom in the denominator.

We can also characterize the distribution of θ_1 in general. Conditional on the factor portfolio returns for the distribution of θ_1 we have:

$$\theta_1 \sim F_{N,T-N-K}(\lambda), \tag{28}$$

$$\lambda = T[1 + \hat{\mu}_p' \hat{\Omega}_p^{-1} \hat{\mu}_p]^{-1} \alpha' \Sigma^{-1} \alpha \tag{29}$$

where λ is the noncentrality parameter of the F distribution.[13]

5.2 Testing Approach.

In this approach we consider the distribution of θ_1 under two different alternatives. The alternatives can be separated by their implications for the maximum value of the squared Sharpe measure. With the risk based multifactor alternative there will be an upper bound on the squared Sharpe measure, whereas with the nonrisk based alternatives the maximum squared Sharpe measure in principle can be unbounded (as the number of assets increases).

First we consider the distribution of θ_1 under the alternative hypothesis when deviations are due to missing factors. Drawing on the results for the squared Sharpe measures, for the noncentrality parameter of the F distribution we have:

$$\lambda = T\,[1 + \hat{\mu}_p' \hat{\Omega}_p^{-1} \hat{\mu}_p]^{-1}\ s_{h_s}^2. \tag{30}$$

From (24), the third term in (30) is bounded above by s_h^2 and positive. The second term is bounded between zero and one. Thus we have an upper bound for λ,

$$\lambda < Ts_h^2 \leq Ts_q^2. \tag{31}$$

The second inequality follows from the fact that the tangency portfolio q has the maximum Sharpe measure of any asset or portfolio.[14]

Given a maximum value for the squared Sharpe measure, the upper bound on the noncentrality parameter can be important. With this bound, independent of how one arranges the assets to be included as dependent variables in the pricing model regression and for any

[13] If $K = 0$ then the term $[1 + \hat{\mu}_p' \hat{\Omega}_p^{-1} \hat{\mu}_p]^{-1}$ will not appear in (27) and in (29) and θ_1 will be unconditionally distributed non-central F.

[14] The first half of this bound appears in MacKinlay (1987) for the case of the Sharpe–Lintner CAPM. Related results appear in Kandel and Stambaugh (1987), Shanken (1987a), and Hansen and Jagannathan (1991).

value of N,[15] there is a limit on the distance between the null distribution and the distribution when the alternative is missing factors.[16] All the assets can be mispriced and yet the bound will still apply. As a consequence one should be cautious in interpreting rejections of the zero-intercept as evidence in favor of a model with more risk factors.

In contrast, when the alternative one has in mind is that the source of nonzero intercepts is nonrisk based such as data snooping, market frictions, or market irrationalities, the notion of a maximum squared Sharpe measure is not useful. The squared Sharpe measure (and the noncentrality parameter) are in principle unbounded. When comparing alternatives with the intercepts of about the same magnitude, in general, we would expect to see larger test statistics in this nonrisk based case.

We can examine the potential informativeness of the above analysis by considering alternatives with realistic parameter values. We construct the distribution of the test statistic for three cases: the null hypothesis, the missing risk factors alternative, and the nonrisk based alternative. For the risk based alternative, we draw on a framework designed to be similar to that in Fama and French (1993). For the nonrisk based alternative we use a setup that is consistent with the analysis of Lo and MacKinlay (1990) and the work of Lakonishok, Shleifer, and Vishny (1993).

We study a one factor asset pricing model using a time series of the excess returns for 32 portfolios for the dependent variable. The one factor (independent variable) is the excess return of the market so that the zero-intercept null hypothesis is the CAPM. The length of the time series is 342 months. This setup corresponds to that of Fama and French (1993) Table 9 regression (ii). For the null distribution of the test statistic θ_1 we have:

$$\theta_1 \quad \sim \quad F_{32,309}(0). \tag{32}$$

To define the distribution of θ_1 under the alternatives of interest we need to specify the parameters necessary to calculate the noncentrality parameter. For the risk based alter-

[15] In practice when using the F-test it will be necessary for N to be less than $T - K$ so that $\hat{\Sigma}$ will be of full rank.

[16] See MacKinlay (1987) for a complete analysis of this point.

native, given a value for the Sharpe measure squared of the optimal orthogonal portfolio, one can consider the distribution corresponding to the upper bound of the corresponding noncentrality parameter. The Sharpe measure of the optimal orthogonal portfolio can be calculated using (19) given the Sharpe measures squared of the tangency portfolio and of the included factor portfolio. My view is that in a perfect capital markets setting *ex ante* a reasonable value for the Sharpe measure squared of the tangency portfolio for an observation interval of one month is 0.031 (or approximately 0.6 for the Sharpe measure on an annualized basis). This value, for example, corresponds to a portfolio with an annual expected excess return of 10% and a standard deviation of 16%. If the maximum squared Sharpe measure of the included factor portfolios is the *ex post* squared Sharpe meansure of the CRSP value weighted index, the implied maximum squared Sharpe measure for the optimal orthogonal porfolio is 0.021. This monthly value of 0.021 would be consistent with a porfolio which has an annualized mean excess return of 8% and annualized standard deviation of 16%.

The selection of the above Sharpe measure can be rationalized both theoretically and empirically. For theoretical justification we can consider the Sharpe measures of equity returns in the literature examining the the equity risk premium puzzle.[17] While the focus of this research does not concern the Sharpe measure, it can be calculated from the analysis provided by Cecchetti and Mark (1990) and Kandel and Stambaugh (1991). Both of these papers are informative for the question at hand since they do not assume any imperfections in the asset markets. If their models with reasonable parameters imply Sharpe measures that are higher than the value selected for use in this paper, one would want to reconsider the selected value. However, one should not completely rely on the measures from these papers for justification. In the presented models the aggregate equity portfolio generally will not be mean variance efficient and therefore need not have the highest Sharpe measure of all equity portfolios.

Common to the papers is the use of a representative agent framework and a Markov switching model for the consumption process. The parameters of the consumption process are chosen to match estimates from the data. Cecchetti and Mark, using the standard

[17]See Mehra and Prescott (1985) for the original discussion of this puzzle.

time-separable constant relative risk aversion utility function, specify range of values for the time preference parameter and the risk aversion coefficient. For each pair of values they generate the implied theoretical unconditional mean and standard deviation of the equity risk premium from which the Sharpe measures can be calculated. The annualized Sharpe measures range from 0.08 to 0.16, substantially below the value of 0.60 suggested above.

Kandel and Stambaugh allow for more general preferences. For the representative agent, a class is used which allows separation of the effects of risk aversion and intertemporal substitution. The standard time-separable model is a special case with the elasticity of intertemporal substitution equal to the inverse of the risk aversion coefficient. They set the monthly rate of time preference at 0.9978 and consider 16 pairs of the risk aversion coefficient and the intertemporal substitution parameter. The risk aversion coefficient varies from 1/2 to 29 and the intertemporal substitution parameter varies from 1/29 to 2. For 13 of the sixteen cases the annual Sharpe measure of equity is less than 0.6. The three cases where the Sharpe measure is greater than 0.6,[18] seem implausible since they imply the equity risk premium and the interest rate have almost the same variance. Historically the variance of the equity risk premium has been substantially higher the *ex post* variance of the real interest rate. In aggregate, the results in these papers are consistent with the value specified for the maximum Sharpe measure squared in the context of frictionless asset markets.

We can also ask what Sharpe measure is empirically reasonable. To do this, we present historical Sharpe measures are presented for a number of broad based indices. These measures, some of which represent portfolios actually held, are reported in Table 1. For each index two estimates are presented, the *ex post* measure (based on maximum likelihood estimates) and an unbiased squared Sharpe measure estimate. For the July 1963 through December 1991 period the squared Sharpe measures are presented for the CRSP value weighted index, the CRSP small stock (tenth decile) portfolio, and the *ex post* optimal portfolio of the two above indices plus the long term government index and the corporate bond index distributed by CRSP in the SBBI file. The small stock portfolio has a monthly squared Sharpe measure of 0.013 (or 0.010 using the unbiased estimate) substantially below the im-

[18] These are the cases with high values for both the risk aversion parameter and the intertemporal substitution parameter.

plied value for the tangency portfolio. The *ex post* optimal four index porfolio's measure is only slightly higher at 0.014.

Table 1 also contains results for the period from January 1981 through June 1992 results for the S&P 500 Index, a growth index and a value index.[19] The source of the return statistics used to calculate the measures is Capaul, Rowley, and Sharpe (1993). These results provide a useful perspective on the maximum magnitudes of Sharpe measures since it is generally acknowledged that the 1980's is a period of strong stock market performance especially for value-based investment strategies. Given this characterization, one would expect these results to provide a high estimate of possible Sharpe measures. We can see that the Sharpe measures from this period are very much in line with (and lower than) the value used in the analysis of the risk based alternative. The highest *ex post* estimate is 0.021 for the value index. Generally, I interpret the evidence in this table as supporting the measure selected to calibrate the analysis for the risk based alternative.

Proceeding using a squared Sharpe measure of 0.021 for the optimal orthogonal portfolio to calculate λ, for the distribution of θ_1 we have:

$$\theta_1 \quad \sim \quad F_{32,309}(7.1). \tag{33}$$

This distribution will be used to characterize the risk based alternative.

I specify the distribution for two nonrisk based alternatives by specifying values of α, Σ, and $\hat{\mu}_p' \hat{\Omega}_p^{-1} \hat{\mu}_p$ and the calculating λ from (29). To specify the intercepts I assume the elements of α are normally distributed with a mean of zero. We consider two values for the standard deviation, 0.0007 and 0.001. When the standard deviation of the elements of α is 0.001 about 95% of the alphas will lie between -0.002 and +0.002, a annualized spread of about 4.8%. A standard deviation of 0.0007 for the alphas would correspond to an annual spread of about 3.4%. These spreads are consistent with spreads that could arise from data-snooping[20] and also plausible and somewhat conservative given the contrarian strategy returns presented in

[19] The growth index contains the S&P 500 stocks with high price to earnings ratios and the value index is constructed from stocks with low price to earnings ratios.

[20] With data-snooping the distribution of θ_1 is not exactly a noncentral F (see Lo and MacKinlay (1990)). However, for the purposes of this paper, the noncentral F will be a good approximation.

Lakonishok, Shleifer, and Vishny. For Σ we use a sample estimate based on portfolios sorted by market capitalization for the period 1963 to 1991 inclusive. The effect of $\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p$ on λ will typically be small, so we set it to zero. To get an idea of a reasonable value for the noncentrality parameter given this alternative, we calculate the expected value of λ given the distributional assumption for the elements of α conditional upon $\Sigma = \hat{\Sigma}$. The expected value of the noncentrality parameter is 39.4 for a standard deviation of 0.0007 and 80.3 for a standard deviation of 0.001. Using these values for the noncentrality parameter of the distribution of θ_1 gives:

$$\theta_1 \sim F_{32,309}(39.4). \tag{34}$$

when $\sigma_\alpha = 0.0007$ and

$$\theta_1 \sim F_{32,309}(80.3) \tag{35}$$

when $\sigma_\alpha = 0.001$.

A plot of the four distributions from (32), (33), (34), and (35) is in Figure 1. The vertical bar on the plot represents the value 1.91 which Fama and French calculate for the test statistic. From this figure notice that the null hypothesis distribution and the risk based alternative distribution are quite close together.[21] This reflects the impact of the upperbound on the noncentrality parameter. In contrast the nonrisk based alternatives' distributions are far to the right of the other two distributions consistent with the noncentrality parameter being unbounded for these alternatives.

What do we learn from this plot? I would claim two things. First, if we want to distinguish among risk based linear asset pricing models the zero-intercept test is not particularly useful because the null distribution and the alternative distribution have substantial overlap. Second, if we want to compare a risk based pricing model with a nonrisk based alternative in mind the zero-intercept test can be very useful since the distributions of the test statistic

[21] See MacKinlay (1987) for detailed analysis of this alternative.

for these alternatives has little overlap.[22]

This analysis can be related to the Fama and French (1993) finding that a model with three factors does a good job in explaining the cross section of expected returns. Because for a given finite cross section under any alternative the inclusion of the optimal orthogonal porfolio will lead to this result, by itself, this result does not support the risk based category. Indeed, the Fama and French approach to building the extra factors will tend to create a portfolio like the optimal orthogonal porfolio independent of the explanation for the CAPM deviations. Their extra factors essentially assign positive weights to the high positive alpha stocks and negative weights to the large negative alpha stocks. This procedure is likely to lead to a portfolio similar to the optimal orthogonal portfolio. This is because the extreme alpha assets are likely to have the largest (in magnitude) weights in the optimal orthogonal portfolio since its weights are proportional to $\Sigma^{\dagger}\alpha$ (see (9)). Further, the fact that, in the Fama and French study when the number of factors is increased to three, the significance of the test statistic only decreases marginally is also consistent with the argument that missing risk factors is not the whole story.

More evidence of the potential importance of nonrisk explanations can be constructed using weekly data. To see why the analysis of weekly data can be informative, consider the biases introduced with market frictions such as the bid–ask spread. Blume and Stambaugh (1983) show that in the presence of the bid–ask bounce, there is an upward bias in the observed returns. For asset i and time period t Blume and Stambaugh show the following approximation for the relation between expected observed returns and expected virtual returns.

$$E(R^{o}_{it}) \;=\; E(R_{it}) \;+\; \vartheta_i \tag{36}$$

where the superscript "o" distinguishes the returns observed with bid-ask bounce contamination from the virtual returns. ϑ_i is the bias which is equal to one fourth of the proportional bid-ask spread squared. The bias will carry over into the intercept of any factor model. Con-

[22]Likelihood analysis provides another interpretation of the plot. Specifically, we can compare the values of the densities for the four alternatives at $\theta_1 = 1.91$. Such a comparison leads to the conclusion that the first nonrisk based alternative is much more likely than the other three.

sider a one factor model where the factor is *ex ante* the tangency portfolio. In this model the intercept for all virtual asset returns will be zero. However, the intercepts for the observed returns and the Sharpe measure squared of the optimal orthogonal portfolio will be nonzero. If the bias of the observed factor return is zero[23] and if the factor return is uncorrelated with the bid-ask bounce process, then for the intercept of the observed returns we will have

$$\alpha_i^o = \vartheta_i \tag{37}$$

since α_i of the virtual return will be 0. Then, for the Sharpe measure squared of the optimal orthogonal portfolio we will have

$$s_h^2 = \vartheta' \Sigma^{o(-1)} \vartheta \tag{38}$$

where Σ^o is the residual covariance matrix for the weekly observed returns and ϑ is the vector of biases for the included portfolios.[24] When the null hypothesis that the intercepts are zero is examined using observed returns violations exist solely due to the presence of the bid-ask spread.

Bias of the type induced by the bid-ask spread is interesting because its magnitude does not depend on the length of the observation interval. As a consequence its effect will statistically be more pronounced with shorter observation intervals when the variance of the virtual returns is smaller. To examine the potential relevance of the above example, the F-test statistic is calculated using a sample of weekly returns for 32 portfolios. The data extends from July 1962 through December 1992 (1591 weeks). NYSE and AMEX stocks are allocated to the portfolios based on beginning of year market capitalization. Each portfolio is allocated an equal number of stocks and the porfolios are equal weighted with rebalancing each week. For these portfolios, using the CRSP value weighted index as the one factor, the F-test statistic is 2.82. Under the null hypothesis, this statistic has a central

[23]This will approximately be the case for a value weighted market index.

[24]The bias of a portfolio will be a weighted average of the bias of the member assets if the weights are independent of the returns process. This will be the case when the portfolio is rebalanced period by period. This will not be the case if the portfolio is weighted to represent a buy and hold strategy (as is a value weighted portfolio). In this latter case the bias at the protfolio level will be minimal.

F distribution with 32 degrees of freedom in the numerator and 1558 degrees of freedom in the denominator.[25] This statistic can be cast in terms of the alternatives presented in Figure 1 since the noncentrality paramter of the F distribution will be approximately invariant to the observation interval and hence only the degrees of freedom need to be adjusted. Figure 2 presents the results which correspond to the weekly observation interval. Basically, these results reinforce the monthly observation results – the observed statistic is most consistent with the nonrisk based category.

In summary, what can be said about the risk based multifactor alternatives versus the nonrisk based arguments for deviations from the Sharpe–Lintner CAPM? The results suggest that the risk based missing risk factors argument is not the whole story. From Figures 1 and 2 we can see that the test statistic is still in the upper tail when we tabulate the distribution of θ_1 in the presence of missing risk factors. The p-value using this distribution is 0.03 for the monthly data and less than 0.001 for weekly data. Hence there is a lack of support for the view that missing factors completely explain the deviations.

On the other hand, given the parametrization considered, there is some support for the nonrisk based alternative views. The test statistic falls almost in the middle of the nonrisk based alternative with the lower standard deviation of the elements of alpha. Several of the nonrisk based alternatives could equally well explain the results. Different nonrisk based views can give the same noncentrality parameter and test statistic distribution. The results are consistent with the data snooping alternative of Lo and MacKinlay (1990), with the related sample selection biases discussed by Kothari, Shanken, and Sloan (1993) and Breen and Korajczyk (1993) and with the presence of market inefficiencies. The bottom line is that the analysis suggests that more than missing risk factors is needed to explain the empirical results.

5.3 Estimation Approach.

In this section we use an estimation approach to make inferences about possible values for Sharpe measures. An estimator for the squared Sharpe measure of the optimal orthogonal

[25]Diagnostics reveal some serial correlation in the residuals of the weekly one factor model. Given this the null distribution will not be exactly central F.

portfolio for a given subset of assets is presented. Using this estimator and its variance, confidence intervals for the squared Sharpe measure can be constructed facilitating judgements on the question of the value implied by the data and reasonable alternatives given this value. An unbiased estimator of the squared Sharpe measure is presented.[26] This estimator corrects for the bias that is introduced by searching over N assets to find the maximum. For the estimator we have:

$$\tilde{s}^2_{h_*} = [\theta_1 - \frac{(T-N-K)}{(T-N-K-2)}]\,[\frac{N(T-N-K-2)}{T(T-N-K)}]\,[1+\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p] \tag{39}$$

$$var(\tilde{s}^2_{h_*}|\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p) = [\frac{2(1+\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p)^2}{T^2}] \times$$

$$[\frac{(N+T[1+\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p]^{-1}s^2_{h_*})^2+(N+2T[1+\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p]^{-1}s^2_{h_*})(T-N-K-2)}{(T-N-K-4)}] \tag{40}$$

Conditional on the factor portfolio returns, the estimator of $s^2_{h_*}$ in (39) is unbiased.

$$E[\tilde{s}^2_{h_*}|\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p] = s^2_{h_*} \tag{41}$$

Recall that when $K=0$ the optimal orthogonal portfolio is the tangency portfolio and hence $s^2_{h_*} = s^2_{q_*}$. The estimator can be applied when $K=0$ by setting $\hat{\mu}_p'\hat{\Omega}_p^{-1}\hat{\mu}_p = 0$.

The estimation approach is illustrated using the above estimator for the Fama and French (1993) portfolios. We consider the case of $K=0$ and therefore we are estimating the maximum squared Sharpe measure from 33 assets – the value-weighted CRSP index, the 25 stock portfolios and 7 bond portfolios. (Recall, with $K=0$, $s^2_{h_*} = s^2_{q_*}$.) The estimator of $s^2_{h_*}$ can be readily calculated, but the variance of $\tilde{s}^2_{h_*}$ cannot since it depends on $s^2_{h_*}$. To calculate the variance we use a consistent estimator, $\tilde{s}^2_{h_*}$, and then asymptotically (as T increases) we

[26]This estimator is derived using the fact that θ_1 is distributed as a non-central F variate. Its moments follow from the moments of the non-central F distribution. Results for the tangency portfolio ($K=0$) are presented by Jobson and Korkie (1980).

have:

$$\hat{s}^2_{h_*} \sim N(s^2_{h_*}, \widehat{var}(\hat{s}^2_{h_*})) \tag{42}$$

Using monthly data from July 1963 through December 1991, the estimate of $s^2_{h_*}$ is 0.092 and the asymptotic standard error is 0.044. Thus using this data set, for a two-sided centered 90% confidence interval we have (0.020, 0.163) and for a one-sided 90% confidence interval we have (0.036, ∞). It is worth noting the upward bias of the ex post maximum squared Sharpe measure as an estimator. For the above case the *ex post* maximum is 0.209 substantially higher than the unbiased estimate of 0.092. The bias is particular severe when N is large (relative to T).

In terms of an annualized Sharpe measure, the two-sided interval corresponds to a lower value of 0.49 and an upper value of 1.40, and the one-sided interval corresponds to a lower value of 0.65. Given that the tangency portfolio and the optimal orthogonal portfolio are the same, we can use this interval to provide an indication of the magnitude of the maximum Sharpe measure needed for a set of factor portfolios to explain the cross section of excess returns of portfolios based on market to book value ratios. Consistent with the CRSP value weighted index being unable to explain the returns, its *ex post* Sharpe measure lies well outside the intervals with an annualized value of 0.33. In general one can use the confidence intervals to decide on promising alternatives. For example, if one believes that *ex ante* Sharpe measures in the 90% confidence interval are unlikely in a risk based world, then the nonrisk based alternatives may provide an attractive area for future study.

6 Asymptotic Arbitrage in Finite Economies.

In the absence of the link between the model deviation and the residual variance expressed in (17) asymptotic arbitrage opportunities can exist. However, the analysis of this paper is based on the importance of the link in a finite economy. It is this importance that will be illustrated in this section. We will use a simple comparison of two economies, economy A where the link is present and economy B where the link is absent. The absence of the link is the only distinguishing feature of economy B. For each economy the behavior of the maximum Sharpe measure squared as a function of the number of securities is examined.

For the analysis specification of the mean excess return vector and the covariance matrix is necessary. We will draw on the previously introduced notation. In addition to the riskfree asset, assume there exists N risky assets with mean excess return μ and nonsingular covariance matrix V, and a risky factor portfolio with mean excess return μ_p and variance σ_p^2. The factor portfolio is not a linear combination of the N assets.[27] For both economies A and B:

$$\mu = \alpha + \beta\mu_p \tag{43}$$

$$V = \beta\beta'\sigma_p^2 + \delta\delta'\sigma_h^2 + I\sigma_\epsilon^2. \tag{44}$$

Given the above mean and covariance matrix and the assumption that the factor portfolio p is a holdable asset, the maximum Sharpe measure squared for economy I is:

$$s_I^2 = s_p^2 + \alpha'(\delta\delta'\sigma_h^2 + I\sigma_\epsilon^2)^{-1}\alpha \tag{45}$$

Analytically inverting $(\delta\delta'\sigma_h^2 + I\sigma_\epsilon^2)$ and simplifying, (45) can be expressed as:

$$s_I^2 = s_p^2 + \frac{1}{\sigma_\epsilon^2}[\alpha'\alpha + \frac{\sigma_h^2(\alpha'\delta)^2}{(\sigma_\epsilon^2 + \sigma_h^2\delta'\delta)}] \tag{46}$$

To complete the specification, the cross-sectional properties of the elements of α and δ are required. We assume the elements of α to be cross-sectionally independent and identically distributed,

$$\alpha_i \sim IID(0, \sigma_\alpha^2) \qquad i = 1, ..., N. \tag{47}$$

The specification of the distribution of the elements of δ conditional on α differentiates

[27]This criterion can be met by eliminating one of the assets which is included in the factor portfolio if necessary.

economies A and B. For economy A

$$\delta_i|\alpha \quad \sim \quad IID(\alpha_i, 0) \qquad i = 1, ..., N, \tag{48}$$

and for economy B

$$\delta_i|\alpha \quad \sim \quad IID(0, \sigma_\alpha^2) \qquad i = 1, ..., N. \tag{49}$$

Unconditionally the cross-sectional distribution of δ will be the same for both economies, but for economy A conditional on α, δ is fixed. This incorporates the deviation - residual variance link. Because δ is independent of α in economy B, the link is absent.

Using (46) and the cross-sectional distributional properties of the elements of α and δ, an approximation for the maximum Sharpe measure squared for each economy can be derived. For both economies $\frac{1}{N}\alpha'\alpha$ converges to σ_α^2, and $\frac{1}{N}\delta'\delta$ converges to σ_α^2. For economy A $\frac{1}{N^2}(\alpha'\delta)^2$ converges to σ_α^4 and for economy B $\frac{1}{N}(\alpha'\delta)^2$ converges to σ_α^4. Substituting these limits into (46) gives approximations of the maximum Sharpe measures squared for each economy.[28] Substitution into (46) for economy A gives

$$s_A^2 \;=\; s_p^2 \;+\; \frac{N\sigma_\alpha^2}{\sigma_\epsilon^2 + N\sigma_h^2\sigma_\alpha^2}, \tag{50}$$

and for economy B gives

$$s_B^2 \;=\; s_p^2 \;+\; N\frac{\sigma_\alpha^2}{\sigma_\epsilon^2}[1 - \frac{\sigma_h^2\sigma_\alpha^2}{\sigma_\epsilon^2 + N\sigma_h^2\sigma_\alpha^2}]. \tag{51}$$

The importance of the link asymptotically can be confirmed by considering the values of s_q^2 in (50) and (51) for large N. For economy A and large N

$$s_q^2 \;=\; s_p^2 \;+\; \frac{1}{\sigma_h^2}, \tag{52}$$

[28]The accuracy of these approximations for values of N equal to 100 and higher was examined. Simulations showed that these approximations are very precise.

and for economy B

$$s_q^2 = s_p^2 + N[\frac{\sigma_\alpha^2}{\sigma_\epsilon^2}]. \qquad (53)$$

The maximum Sharpe measure squared is bounded as N increases for economy A and unbounded for economy B. Using the correspondence between boundedness of the maximum Sharpe measure squared and the absence of asymptotic opportunities (see Ingersoll (1984) Theorem I) there will be asymptotic arbitrage opportunities only in economy B.

However, our interest here is to examine the importance of the deviation - residual variance link given a finite number of assets. We do this by considering the value of the maximum Sharpe measures for various values of N.[29] The values of N considered are 100, 500, 1000, and 5000. For completeness we also report the maximum Sharpe measure squared for $N = \infty$. Given (50) and (51), to complete the calculations, s_p^2, σ_h^2, σ_ϵ^2 and σ_α^2 must be specified. The parameters are selected so that μ and V are realistic for stock returns measures at a monthy observation interval. The selected parameter values are $s_p^2 = 0.01$, $\sigma_h = 2.66$, $\sigma_\epsilon = 0.05$. Two values are considered for σ_α, 0.001 and 0.002. The results are reported in Table 2. The difference in the behavior of the maximum squared Sharpe measures between economies A and B is dramatic. For economy A the boundedness is apparent as the maximum Sharpe measure squared ranges from 0.023 to 0.030 as N increases from 100 to infinity. For economy A the impact of increasing the cross-sectional variation in the mean return is minimal. Comparing the $\sigma_\alpha = 0.001$ panel to the $\sigma_\alpha = 0.002$ reveals few differences with the exception of differences for the $N = 100$ case. For economy B it is a different story. The maximum Sharpe measure squared is very sensitive to both increasing the number of securities and to increasing the cross-sectional variation in the mean return. For $\sigma_\alpha = 0.002$ the maximum Sharpe measure squared increases from 0.169 to 1.608 as N increases from 100 to 1000. When σ_α increases from 0.001 to 0.002 the maximum Sharpe measure squared increases from 0.21 to 0.80 for N equal to 500.

In addition to the maximum Sharpe measures squared, Table 2 reports the approximate

[29]Shanken (1992) presents related results for an economy similar to B with δ restricted to be 0 for $N = 3000$ and $N = 3$ *million*. He notes that for $N = 3$ *million* "something close to a 'pure' arbitrage is possible."

probability that the annual excess return of the maximum Sharpe measure squared portfolio is negative. For this probability calculation, it is assumed that returns are jointly normally distributed and that the mean excess return of the maximum Sharpe measure squared portfolio is non-negative. The mean and variance are annualized by multiplying the monthly values by 12. This probability allows for an economic interpretation of the size of the Sharpe measure. Since the excess return represents a payoff on a zero investment position, if the probability of a negative outcome is 0 then there is an arbitrage opportunity. For economy A this probability is about 28% and stable as N increases. However, for economy B the probability of a negative annual excess return quickly approaches 0. For example, for the case of σ_α equal to 0.002 and N equal to 500 the probability of a negative outcome is less than 0.001.[30] Since negative outcomes can occur, the excess return distributions cannot be completely ruled out on economic grounds. However, in aggregate it appears that, given the above model for economy B, unrealistic investment opportunites can be constructed with a relatively small number of stocks. This is not the case for economy A. The bottom line is that is a perfect capital markets environment, the link between the model deviations and the residual variance is important even with a limited number of securities. Analysis which does not recognize this link is unlikely to shed light on the potential for omitted risk factors to explain the deviations.

7 Conclusion.

Empirical work in economics in general and in finance in particular is *ex post* in nature. Given this, it is often difficult to discriminate between various explanations for observed phenomena. A partial solution to this difficulty is to examine the alternatives and make judgements from an *ex ante* point of view. The current explanations of asset pricing empirical results are particularly well suited to *ex ante* analysis. This paper presents a framework based on the economics of mean-variance analysis to address and reinterpret prior empirical results.

[30]To put these probabilities of negative annual excess returns into perspective, they can be compared to actual results for the S&P Index and the CRSP small stock index. For the 67 years from 1926 through 1992 the excess return of the S&P index has been negative 37.3% of the years and the excess return of the CRSP small stock index has been negative 34.3% of the years. These percentages are not particularly sensitive to the time period considered. Over the 30 year period from 1963 through 1992 the S&P Index has been negative 36.7% of the time and the small stock index has been negative 30.0% of the time.

It has been common to look to multifactor asset pricing models as an alternative to the Sharpe-Lintner CAPM. However, the results in this paper suggest that looking at other alternatives may be fruitful. The evidence against the CAPM can also be interpreted as evidence against multifactor alternatives being the complete explanation. It is improbable that multifactor models on their own will explain the deviations from the CAPM. Generally, the results suggest that more can be learned by considering the likelihood of various existing empirical results under differing specific economic models.

References

Amihud, Y., and H. Mendleson, 1986, "Asset pricing and the bid-ask spread," *Journal of Financial Economics* 17, 223–250.

Breen, W., and R. Korajczyk, 1993, "On selection biases in book-to-market based tests of asset pricing models," Working Paper #167, Northwestern University.

Campbell, J., 1987, "Stock returns and the term structure," *Journal of Financial Economics* 18, 373–399.

Capaul, C., Rowley, I. and W. Sharpe, 1993, "International value and growth stock returns," *Financial Analysts Journal* January–February, 27–36.

Cecchetti, S., and N. Mark, 1990, "Evaluating empirical tests of asset pricing models," *American Economic Review* 80, 48–51.

Chamberlain, G. and M. Rothschild, 1983, "Arbitrage, factor structure, and mean–variance analysis on large asset markets," *Econometrica* 51, 1281–1304.

Chan, Louis K. C. and J. Lakonishok, 1993, "Are reports of beta's death premature?" *Journal of Portfolio Management* 19, 51–62.

Connor, G. and R. Korajczyk, 1988, "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics* 21, 255–290.

Conrad, J., and G. Kaul, 1993, "Long-term market overreaction or biases in computed returns?" *Journal of Finance* 48, 39–63.

DeBondt, W., and R. Thaler, 1985, "Does the stock market overreact?" *Journal of Finance* 40, 793–805.

Fama, E., 1993, "Multifactor portfolio efficiency and multifactor asset pricing models," CRSP Working Paper, University of Chicago.

Fama, E. and K. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics* 33, 3–56.

Gibbons, M. R., 1982, "Multivariate tests of financial models: A new approach," *Journal of Financial Economics* 10, 3–27.

Gibbons, M., Ross, S. and J. Shanken, 1989, "A test of the efficiency of a given portfolio," *Econometrica* 57, 1121–1152.

Hansen, L., and R. Jagannathan, 1991, "Implications of security market data for models of dynamic economies," *Journal of Political Economy* 99, 225–262.

Hansen, L., and R. Jagannathan, 1994, "Assessing specification errors in stochastic discount factor models," working paper, University of Minnesota.

Huberman, G. Kandel, S. and R. Stambaugh, 1987, "Mimicking portfolios and exact arbitrage pricing," *Journal of Finance* 42, 1–10.

Ingersoll, J., 1984, "Some results in the theory of arbitrage pricing," *Journal of Finance* 39, 1021–1039.

Jobson, D., and R. Korkie, 1980, "Estimation for Markowitz efficient portfolios," *Journal of the American Statistical Association* 75, 544–554.

Jobson, D., and R. Korkie, 1982, "Potential performance and tests of portfolio efficiency," *Journal of Financial Economics* 10, 433–466.

Kandel, S. and R. Stambaugh, 1987, "On correlations and inferences about mean-variance efficiency," *Journal of Financial Economics* 18, 61–90.

Kandel, S. and R. Stambaugh, 1990, "A mean-variance framework for tests of asset pricing models," *Review of Financial Studies* 2, 125–156.

Kandel, S. and R. Stambaugh, 1991, "Asset returns and intertemporal preferences," *Journal of Monetary Economics* 27, 39–71.

Kandel, S. and R. Stambaugh, 1994, "Portfolio inefficiency and the cross-section of mean returns," Rodney White Center for Financial Research Paper, Wharton School, University of Pennsylvania.

Kothari, S., J. Shanken, and R. Sloan, 1993, "Another look at the cross-section of expected returns," manuscript, Wharton School, University of Pennsylvania.

Lakonishok, J., A. Shleifer, and R. Vishny, 1993, "Contrarian investment, extrapolation, and risk," unpublished working paper, University of Illinois, University of Chicago, and Harvard University.

Lehmann, B., 1987, "Orthogonal frontiers and alternative mean variance efficiency tests," *Journal of Finance* 42, 601–619.

Lehmann, B., 1989, "Mean-variance efficiency tests in large cross-sections," working paper, Columbia University and National Bureau of Economic Research.

Lehmann, B., 1992, "Empirical testing of asset pricing models," in *The New Palgrave Dictionary of Money and Finance*, Peter Newman, Murray Milgate, and John Eatwell eds., Stockton Press (New York), 749–759.

Lehmann, B. N. and D. Modest, 1988, "The empirical foundations of the arbitrage pricing theory," *Journal of Financial Economics* 21, 213–254.

Lo, A. and A. C. MacKinlay, 1990, "Data-snooping biases in tests of financial asset pricing models," *Review of Financial Studies* 3, 431–467.

Luttmer, E., 1991, "Asset pricing in economies with frictions," working paper, University of Chicago, November.

MacKinlay, A. C., 1987, "On multivariate tests of the CAPM," *Journal of Financial Economics* 18, 341–372.

MacKinlay, A. C., and M. Richardson, 1991, "Using generalized methods of moments to test mean-variance efficiency," *Journal of Finance* 46, 511–527.

Merton, R., 1973, "An intertemporal capital asset pricing model," *Econometrica* 41, 867–887.

Roll, R., 1977, "A critique of the asset pricing theory's tests: Part I," *Journal of Financial Economics* 4, 129–176.

Roll, R., 1980, "Orthogonal portfolios," *Journal of Financial and Quantitative Analysis* 15, 1005–1023.

Roll, R., and S. Ross, 1994, "On the cross-sectional relation between expected returns and betas," *Journal of Finance* 49, 101–122.

Ross, S., 1976, "The arbitrage theory of capital asset pricing," *Journal of Economic Theory* 13, 341–360.

Shanken, J., 1985, "Multivariate tests of the zero-beta CAPM," *Journal of Financial Economics* 14, 327–348.

Shanken, J., 1987a, "Multivariate proxies and asset pricing relations: living with the Roll critique," *Journal of Financial Economics* 18, 91–110.

Shanken, J., 1992, "The current state of the arbitrage pricing theory," *Journal of Finance* 47, 1569–1574.

Stambaugh, R. F., 1982, "On the exclusion of assets from tests of the two parameter model," *Journal of Financial Economics* 10, 235–268.

Table 1

Historical Sharpe measures for selected stock indices. $\hat{s}_h^2$ is the monthly expost Sharpe measure squared and $\hat{s}_h(ann)$ is the positive square root of this measure annualized. $\tilde{s}_h^2$ is an unbiased estimate of the monthly Sharpe measure squared and $\tilde{s}_h(ann)$ is the positive square root of this measure annualized. The portfolio of four indices is the portfolio with the maximum expost Sharpe measure squared . The four indices are CRSP value-weighted index, CRSP small stock decile, CRSP longterm government bond index and CRSP corporate bond index. The bond indices are from the CRSP SBBI file. All indices are based on total return.

Time Period	Index	$\hat{s}_h^2$	$\hat{s}_h(ann)$	$\tilde{s}_h^2$	$\tilde{s}_h(ann)$
6307 - 9112	CRSP VW Index	0.0091	0.33	0.0061	0.27
6307 - 9112	Small Stock Decile	0.0142	0.40	0.0100	0.35
6307 - 9112	Portfolio of four Indices	0.0145	0.41	0.0021	0.16
8101 - 9206	S&P 500 Index	0.0161	0.44	0.0085	0.32
8101 - 9206	S&P-BARRA Value Index	0.0208	0.50	0.0130	0.40
8101 - 9206	S&P-BARRA Growth Index	0.0108	0.36	0.0033	0.20

Table 2

A comparison of the maximum Sharpe measure squared for two economies denoted A and B. The excess return covariance matrix for the two economies is identical and the cross-sectional dispersion in mean excess returns is identical. The economies differ in that economy A displays stronger dependence between the mean excess returns and the covariance matrix of excess returns. The mean and covariance matrix parameters for the economies are calibrated to correspond roughly to monthly returns – see the text for details. N is the number of securities, s_I^2 is the maximum Sharpe measure squared for economy I, $I = A, B$, and $p(z_I < 0)$ is the approximate probability for economy I that the annual return of the maximum Sharpe measure squared portfolio will be less than the risk free return assuming that monthly returns are jointly normally distributed and that the mean excess return is positive. σ_α is the cross-sectional standard deviation of the component of the mean return that is explained by a second factor in economy A and that is not explained by a common factor in economy B.

N	s_A^2	$p(z_A < 0)$	s_B^2	$p(z_B < 0)$
$\sigma_\alpha = 0.001$				
100	0.023	0.298	0.050	0.220
500	0.028	0.280	0.210	0.056
1000	0.029	0.277	0.410	0.013
5000	0.030	0.275	2.010	**
∞	0.030	0.274	∞	**
$\sigma_\alpha = 0.002$				
100	0.028	0.282	0.169	0.077
500	0.030	0.276	0.808	**
1000	0.030	0.275	1.608	**
5000	0.030	0.274	8.008	**
∞	0.030	0.274	∞	**

** Less than 0.001.

Figure 1

Distributions for the CAPM zero-intercept test statistic for four alternatives. Alternative 1 is the CAPM (null hypothesis). Alternative 2 is the risk based alternative (deviations from the CAPM are from missing risk factors). Alternatives 3 and 4 are the nonrisk based alternative (deviations from the CAPM are unrelated to risk). The distributions are $F_{32,309}(0)$, $F_{32,309}(7.1)$, $F_{32,309}(39.4)$, and $F_{32,309}(80.3)$ for alternatives 1, 2, 3, and 4 respectively. The degrees of freedom are set to correspond to monthly observations from July 1963 to December 1992 (342 observations). Using 25 stock portfolios and 7 bond portfolios, and the CRSP value-weighted index as proxy for the market portfolio the test statistic is 1.91.

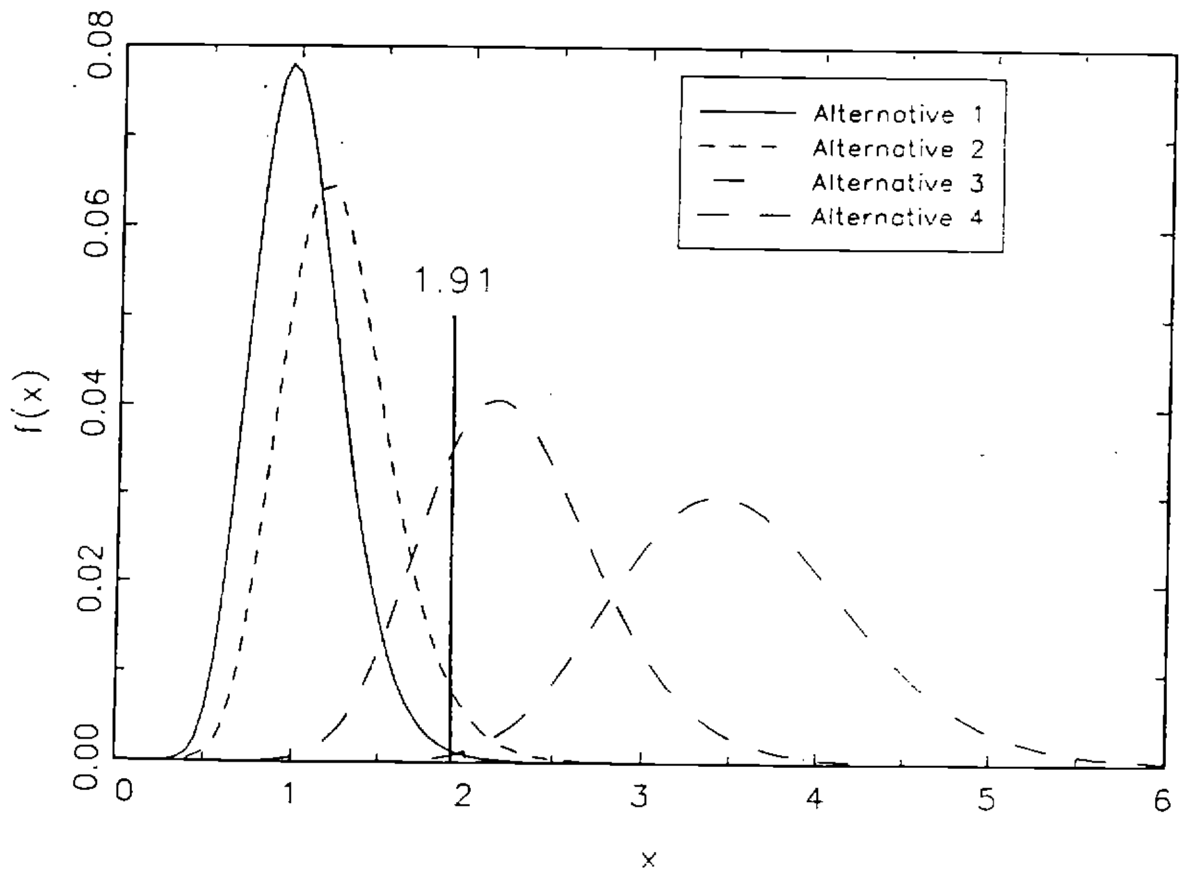

Figure 2

Distributions for the CAPM zero-intercept test statistic for four alternatives. Alternative 1 is the CAPM (null hypothesis). Alternative 2 is the risk based alternative (deviations from the CAPM are from missing risk factors). Alternatives 3 and 4 are the nonrisk based alternative (deviations from the CAPM are unrelated to risk). The distributions are $F_{32,1558}(0)$, $F_{32,1558}(7.1)$, $F_{32,1558}(39.4)$, and $F_{32,1558}(80.3)$ for alternatives 1, 2, 3, and 4 respectively. The degrees of freedom are set to correspond to weekly observations from July 1963 to December 1992 (1591 observations). Using 32 stock portfolios and the CRSP value-weighted index as a proxy for the market portfolio the test statistic is 2.82.