NBER WORKING PAPER SERIES

MUTUAL FUND PERFORMANCE EVALUATION: A COMPARISON OF BENCHMARKS AND BENCHMARK COMPARISONS

Bruce N. Lehmann

David M. Modest

Working Paper No. 1721

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1985

We would like to express our gratitude to Roy Henriksson for providing us with his mutual fund data base, Karen Bettauer for assistance in updating it, and the Faculty Research Fund of the Columbia Business School and the Institute for Quantitative Research in Finance for their support. We are also grateful for comments from Allan Kleidon, Robert Korajozyk, Arthur Warga and seminar participants at Stanford University, and the Universities of California at Berkeley and Los Angeles. The research reported here is part of the NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #1721
October 1985

Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons

ABSTRACT

Our primary goal in this paper is to ascertain whether the absolute and relative rankings of managed funds are sensitive to the benchmark chosen to measure normal performance. We employ the standard CAPM benchmarks and a variety of APT benchmarks to investigate this question. We found that there is little similarity between the absolute and relative mutual fund rankings obtained from alternative benchmarks which suggests the importance of knowing the appropriate model for risk and expected return in this context. In addition, the rankings are quite sensitive to the method used to construct the APT benchmark. One would reach very different conclusions about the funds' performance using smaller numbers of securities in the analysis or the less efficient methods for estimating the necessary factor models than one would arrive at using the maximum likelihood procedures with 750 securities. We did, however, find the rankings of the funds are not very sensitive to the exact number of common sources of systematic risk that are assumed to impinge on security returns. Finally, we found statistically significant measured abnormal performance using all the benchmarks. The economic explanation of this phenomenon appears to be an open question.

Bruce Lehmann
Graduate School of Business
Columbia University
New York, NY 10027

David M. Modest
Graduate School of Business
Columbia University
New York, NY 10027

I. Introduction

Thirty years of scientific progress in financial economics has left the central problem of portfolio performance evaluation largely unresolved. This unhappy state of affairs persists despite broad agreement on an intuitive level that an actively managed portfolio with superior performance should exhibit higher returns on average than a passively managed portolio with the same amount of risk. Unfortunately, two obstacles stand in the way of implementing this intuitive notion of superior performance to evaluate the track record of managed funds. The first difficulty stems from disagreement on the appropriate way to quantify risk and hence on what constitutes normal performance. The second problem concerns errors in inference that can arise when portfolio managers can, in fact, outperform the market.

In order to measure abnormal performance by mutual funds, it is necessary to have a benchmark for normal performance. Modern portfolio theory purports to provide such a standard of comparison — that combination of the market portfolio and the riskless asset which is of comparable risk. Not surprisingly, numerous investigators have employed the Capital Asset Pricing Model (**CAPM**) to evaluate the performance of mutual funds.[1] Roll(1977,1978), however, has forcefully argued that the use of the **CAPM** as a benchmark in performance evaluation is logically inconsistent under the assumptions of the model since any measured abnormal performance can only occur when the market proxy is inefficient. In the absence of any systematic evidence of abnormal performance by mutual funds, Roll's critique would appear to be an academic point. Yet there is plenty of ancillary empirical evidence indicating the mean-variance inefficiency of the usual indices, including the anomalies involving dividend yield, firm size, and price/earnings ratios, which leads one to question the use of the usual **CAPM** market proxies as an appropriate benchmark.[2]

[1] These include Treynor(1965), Treynor and Mazuy(1966), Sharpe(1966), Jensen(1968,-1969), Friend, Blume, and Crockett(1970), McDonald(1973,1974), and Mains(1977).

[2] See, for example, Cannistraro(1973), Basu(1977), Litzenberger and Ramaswamy(1979), Banz(1981), Reinganum(1981). The direct evidence on the validity of the **CAPM** is mixed. Studies examining the Sharpe-Lintner version of the model such as Black,Jensen and Scholes(1972), Fama and MacBeth(1973), and Blume and Friend(1973) have rejected the hypothesis that the usual indices are the tangency portfolios associated with the riskless rate. The evidence on the zero-beta **CAPM** also has resulted in mixed conclusions concerning the mean-variance efficiency of these indices as evidenced by the results of Gibbons(1982), Jobson and Korkie(1982), Shanken(1984), and Stambaugh(1982).

The apparent inefficiency of the usual market proxies coupled with concern over the testability of the **CAPM** (e.g. Roll(1977)) has led researchers to explore alternative theories of asset pricing. One theory which has stimulated much recent research is the Arbitrage Pricing Theory(**APT**) developed by Ross(1976,1977). In that seminal work, Ross contended that the fundamental idea imbedded in the **CAPM** is the differentiation between idiosyncratic risk that can be eliminated in large portfolios through diversification and systematic risk that is pervasive and cannot be easily avoided.[3] He further reasoned that systematic risk need not be adequately represented by a single common factor such as the return on the market and instead presumed that there are **K** common sources of covariation (risk) affecting security returns. These **K** factors constitute another potential benchmark with which to measure normal performance.

Since the theory does not require that these sources of systematic risk be specified *a priori*, empirical implementation of the **APT** usually involves the construction of basis or reference portfolios to mimic the factors. There are any number of ways to form such portfolios. There are not only several methods for forming the portfolios but there are different procedures for estimating the factor models of security returns that underly these computations. In addition, one has discretion over how many securities to use in the analysis and how many factors to extract. In principle, each variant provides another potential benchmark.

Even if there were no question about the appropriate benchmark, it is still difficult to measure managerial performance when mutual fund managers are superior investors. This second difficulty arises from problems associated with measuring portfolio risk when managers act on private information and, as a consequence, continually revise the composition of their portfolios and the risk level of their funds. After relaxing the assumptions of the **CAPM**, Mayers and Rice(1979) provide sufficient conditions under which the standard Security Market Line analysis is a valid measure of portfolio performance ability. Unfortunately, as shown by Dybvig and Ross(1981), their sanguine conclusion rests on the assumption that managers possess no market timing ability and that any abnormal performance is due to stock selection. This occurs because uninformed investors, unable

[3] The idea that the important distinction is between diversifiable and nondiversifiable is also captured in the single index market model of security returns introduced by Markowitz(1952) and developed and extended by Sharpe(1963,1967).

to observe managers' private information signals or actual portfolio choices, may perceive implicit changes in expected returns due to market timing as needless additions to variance when they are forced to draw inferences solely on the basis of the realized returns of the portfolio. In this case, the usual Security Market Line analysis will detect abnormal performance but will be unable to distinguish superior from inferior ability.[4]

Our primary goal in this paper is to ascertain whether the absolute and relative rankings of managed funds are sensitive to the benchmark chosen to measure normal performance. An ancillary goal of the paper is to examine the efficacy of Security Market Line type analysis in the evaluation of mutual fund performance given the shifting composition of managed portfolios. Not only do we compare and contrast the **CAPM** and the **APT**, but we also examine the different basis portfolio construction methods that have been suggested in the literature to produce portfolios that are highly correlated with the common factors underlying the **APT**. The question of whether comparatively inexpensive but statistically inefficient basis portfolio formation procedures lead to quantitatively serious benchmark errors is of special interest. In Lehmann and Modest(1985b) we studied the different procedures suggested to mimic the factors and found evidence that suggests that inexpensive basis portfolio procedures sacrifice a significant amount of statistical precision and seem to do a relatively poor job of mimicking the factors. However, the analysis there left open the question of whether this statistical evidence would translate into meaningful economic differences. Mutual fund performance evaluation provides a natural laboratory for the investigation of these questions.

It is worth emphasizing that previous research would suggest that we should expect few substantive differences in the performance measures implied by alternative risk adjustment procedures. For example, Stambaugh(1983) found that the choice of a market proxy made little difference in **CAPM** tests. Moreover, Roll(1979) found that three market proxies provided nearly identical performance measures for randomly selected portfolios and that these risk adjustment methods produced almost the same rankings as no adjustment at all. Similarly, Copeland and Mayers(1982) and Chen, Copeland, and Mayers(1983) found that the choice of a performance benchmark did not affect inferences regarding the Value Line

[4] The problem of measuring the quality of investment ability has also recently been addressed by Verrechia(1980), Pfleiderer and Bhattacharya(1983), Admati and Ross (1985), and Grinblatt and Titman(1985).

enigma. It is certainly of independent interest to know whether alternative risk measurement procedures yield similar rankings in this context as well.

The application of the **APT** involves numerous technical and economic questions. The next section discusses some of the economic and statistical issues associated with the employment of the **APT** in benchmark comparisons. In Section III we discuss the ability of Jensen(1968,69)-style Security Market Line regressions[5] and Jensen(1972) and Pfleiderer and Bhattacharya(1983) quadratic regressions to detect abnormal performance and market timing ability in the **APT** context. The data is discussed in Section IV. Section V reports the empirical results. We begin by presenting summary statistics concerning the comparative performance of different **APT** benchmarks. These statistics contrast the inferences concerning abnormal mutual fund performance that are arrived at using reference portfolios constructed from different sized cross-sections of securities and alternative estimation methods. We also consider the dependence of the absolute and relative rankings of the on the number of factors which are extracted from the data. The focus then shifts to an examination of the performance of the **CAPM** and **APT** benchmarks. In particular, we examine whether the **APT** has anything different to say about the performance of mutual funds than the **CAPM**. Finally, we use quadratic regressions to examine the problems associated with the shifting composition and risk of managed portfolios. The final section provides concluding remarks concerning the abnormal performance of mutual funds and the comparative merits of alternative benchmarks.

II. Implementing The APT

A. The Arbitrage Pricing Theory

The cornerstone of the **APT** is the statistical assumption that security returns depend on **K** common factors, whose risk cannot be eliminated in *arbitrary* well-diversified portfolios, plus some idiosyncratic risk that (as the number of securities becomes infinite) can be diversified away in such portfolios. Under the assumption that these factors affect securities returns linearly, the factor model for returns takes the form:

[5] This question has also been recently (and independently) examined by Connor and Korajczyk(1984).

$$\tilde{R}_{it} = E_i + \sum_{k=1}^{K} b_{ik}\tilde{\delta}_{kt} + \tilde{\epsilon}_{it} \tag{1}$$

$$\mathbf{E}[\tilde{\delta}_{kt}] = \mathbf{E}[\tilde{\epsilon}_{it}|\delta_{kt}] = 0$$

where:

$\tilde{R}_{it} \equiv$ Return on security i between time t-1 and time t for i=1,...,N

$E_i \equiv$ Expected return on security i

$\tilde{\delta}_{kt} \equiv$ Value taken by the k^{th} common factor { i.e source of systematic risk } between time $t-1$ and t

$b_{ik} \equiv$ sensitivity of the return of security i to the k^{th} common factor { called the factor loading } and

$\tilde{\epsilon}_{it} \equiv$ the idiosyncratic or residual risk of the return on the i^{th} security between time $t-1$ and time t which has zero mean, finite variance and is sufficiently independent across securities for a law of large numbers to apply.

In these circumstances, Ross(1976,1977) showed that it is possible to form zero net investment arbitrage portfolios with no systematic risk and negligible idiosyncratic risk as well. Intuitively, these nearly riskless portfolios should have zero profits in the absence of taxes and transaction costs. Ross formalized this intuition and proved that if the number of securities satisfying the return generating process (1) is large, then, to ensure that riskless arbitrage profits cannot be made, expected returns must satisfy (*approximately*):

$$E_i \approx \lambda_0 + b_{i1}\lambda_1 + ... + b_{ik}\lambda_k \tag{2}$$

where:

$\lambda_0 \equiv$ the intercept in the pricing relation and

$\lambda_k \equiv$ the risk premium on the k^{th} common factor, $k = 1, \ldots, K$.

If the return on the market was the only common factor, then λ_1 would be the expected excess return on the market.

The theory has received considerable attention and has been discussed and further developed in numerous papers including Huberman(1982), Chamberlain and Rothschild-(1983), Chamberlain(1983), Dybvig(1983), Grinblatt and Titman(1983,1984), Chen and

Ingersoll(1983), Connor(1984,1985), and Ingersoll(1984).[6] These extensions have been primarily devoted to three topics: (a) characterization and weakening of the sufficient conditions for (2) to hold as an approximation in large economies; (b) derivation of sufficient conditions for (2) to hold as an equality in large economies; and (c) computation of explicit bounds on the deviations from (2) in finite economies.

For empirical purposes, it is obviously advantageous to be able to assume that equation (2) provides an exact, rather than an approximate, theory of expected returns. Therefore we shall presume sufficient structure so that expected returns exactly satisfy:

$$\underline{E} = \underline{\iota}\lambda_0 + B\underline{\lambda} \tag{3}$$

where:

$\underline{E} \equiv$ the $N \times 1$ vector of expected returns on the N securities,

$\underline{\lambda} \equiv$ the $K \times 1$ vector of risk premia on the **K** factors,

$B \equiv$ the $N \times K$ matrix of individual security factor loadings: b_{ij}, $\quad i = 1, \ldots, N \quad j = 1, \ldots, K$,

and λ_0 is as defined above. Sufficient conditions are given in Grinblatt and Titman(1983), Connor(1984), and Ingersoll(1984). They basically involve the assumptions that investors are not too risk averse, the idiosyncratic risk of the individual assets is not too substantial, and the value of any asset as a proportion of total wealth is not too large. Limiting necessary and sufficient conditions are given in Chamberlain and Rothschild(1983). They show that in large economies where large subsets of security returns follow a factor model, an exact pricing relationship will hold if and only if (in the limit) there is a risky, well-diversified portfolio on the mean-variance efficient frontier constructed from the (countably infinite) subset of returns satisfying the approximate factor structure.

Finally, we differentiate two versions of the **APT** which involve different interpretations of λ_0. Like the **CAPM**, the **APT** has both a riskless rate and a zero beta formulation. The riskless rate version is appropriate when it is possible to form a positive investment portfolio of risky assets whose return variance goes to zero as the number of assets which

[6] For a critique of the theory see Shanken(1982) and Dhrymes,Friend and Gultekin (1984). For replies to these articles, see Dybvig and Ross(1983), Pfleiderer and Reiss(1983), and Roll and Ross(1984).

satisfy an approximate factor structure grows large. In this version of the **APT**, λ_0 is the riskless rate. The zero beta formulation arises when it is not possible to form a limiting riskless portfolio of risky assets. Under this formulation, one of the factors corresponds to the zero beta return. What makes the zero beta factor different from the other factors is that, under an appropriate transformation of the factor space, all securities will have equal sensitivity to it. Hence λ_0 is zero when the zero beta formulation is appropriate.[7] In what follows, we will use both the riskless rate and zero beta formulations of the **APT** to examine whether conclusions regarding the comparative performance of alternative **APT** benchmarks is sensitive to which version of the **APT** is apposite.[8]

B. The APT and Estimation of the Factors

To be sure, testing and using this theory would be straightforward if the common factors $\tilde{\delta}_{kt}$ could be easily identified with observable economic or financial data. Unfortunately, financial theory seems to be capable of rationalizing a wide variety of potential sources of systematic risk. In consequence, several authors have used the statistical method of factor analysis in order to ascertain whether (1) is an appropriate model for security returns and (3) provides an accurate model of expected returns. These studies include Gehr(1975), Roll and Ross(1980), Gibbons(1983), Reinganum(1981), Hughes(1982), Brown and Weinstein(1983), Chen(1983), Dhrymes, Friend and Gultekin(1984,1985), and Lehmann and Modest(1985a).[9]

The first step in obtaining an **APT** benchmark is to construct portfolios which reflect the behavior of the K *unobservable* common factors. There are a number of different portfolio formation procedures that can be used to construct these mimicking portfolios. Four different methods were compared in Lehmann and Modest(1985b): a generalized least squares (GLS) procedure, a variant of the GLS procedure that produces what we call mini-

[7] Note that the distinction between the two versions of the **APT** arises from the feasibility of forming riskless portfolios of risky assets and not from constraints on riskless borrowing and lending or short sales, a point emphasized by Ingersoll(1984). The zero beta version of the **APT** thus differs in a fundamental way from the zero beta version of the **CAPM**.

[8] In Lehmann and Modest(1985a) we provide detailed evidence on the merits of the two versions.

[9] Studies which have investigated factor models of asset returns without reference to arbitrage pricing theory include King(1966), Farrell(1974), Feeney and Hester(1967), Fama and MacBeth(1973), Rosenberg and Marathe(1979), and Arnott(1980).

mum idiosyncratic risk portfolios, and two quadratic programming procedures that impose constraints to produce minimum idiosyncratic risk portfolios with small portfolio weights. The evidence presented there suggests the superiority of the minimum idiosyncratic risk procedure and consequently we used that portfolio formation procedure here. The novelty of the minimum idiosyncratic procedure compared with the more familiar cross-sectional regression methods prompts the following detailed examination.

Since the minimum idiosyncratic risk procedure is a variant of the GLS procedure, it is useful to first consider the statistical intuition underlying generalized least squares methods. Under the assumption that the returns of the N securities under consideration are generated by a **K** factor linear structure as given in (1), we can write the assumed joint return generating process of the N securities as:

$$\underline{\tilde{R}}_t = \underline{E} + B\underline{\tilde{\delta}}_t + \underline{\tilde{\epsilon}}_t \tag{4}$$

where the residual risks $\underline{\tilde{\epsilon}}_t$ are assumed to satisfy:

$$\begin{aligned} \mathbf{E}[\underline{\tilde{\epsilon}}_t|\underline{\tilde{\delta}}_t] &= \underline{0} \\ \mathbf{E}[\underline{\tilde{\epsilon}}_t\underline{\tilde{\epsilon}}_t'|\underline{\tilde{\delta}}_t] &= \Omega \end{aligned} \tag{5}$$

and Ω is a positive definite symmetric matrix. Since the factors are unobservable, the model is not identified without further *a priori* restrictions. We therefore assume that the random factors $\underline{\tilde{\delta}}_t$ (a $K \times 1$ vector) and the corresponding elements of the factor loading matrix, B $(N \times K)$, have been normalized so that:[10]

$$\begin{aligned} \mathbf{E}[\underline{\tilde{\delta}}_t] &= \underline{0} \\ \mathbf{E}[\underline{\tilde{\delta}}_t\underline{\tilde{\delta}}_t'] &= I \end{aligned} \tag{6}$$

Treating the factor loadings in equation (4) as explanatory variables that are measured without error and the common factors as parameters to be estimated, one natural way to proceed is to run an ordinary least squares cross-sectional regression of the excess returns

[10] The elements of B are still not uniquely determined since for all orthogonal matrices T, any matrix $B^* = BT$ will yield the same return generating process. We assume that the necessary $K(K-1)/2$ constraints required to ensure that $T = I$ have been imposed arbitrarily. For example, it is conventional in factor analysis to require $B'\Omega^{-1}B$ to be a diagonal matrix.

of the individual securities, $\underline{\tilde{R}}_t - \underline{E}$, on the factor loadings along the lines of the procedure followed by Fama and MacBeth(1973). The ordinary least squares estimates of the factors at date t would then be given by:[11]

$$\hat{\underline{\delta}}_t^{OLS} = (B'B)^{-1}B'\underline{\tilde{R}}_t \tag{7}$$

Note that the projection matrix $(B'B)^{-1}B'$ can be thought of as the transpose of a (time invariant) $N \times K$ matrix of portfolio weights that can be used in conjunction with the returns $\underline{\tilde{R}}_t$ at any date t to obtain an estimate of the realization of the **K** common factors $\underline{\delta}_t$. The ordinary least squares estimator is not efficient, however, since it ignores the information in the covariance matrix of the residual risks, Ω. A more efficient procedure is the generalized least squares estimator of $\underline{\tilde{\delta}}_t$ given by: [12]

$$\hat{\underline{\delta}}_t^{GLS} = (B'\Omega^{-1}B)^{-1}B'\Omega^{-1}[\underline{\tilde{R}}_t - \underline{E}] \tag{8}$$

This estimator is the minimum variance linear unbiased estimator of $\underline{\tilde{\delta}}_t$ and, in addition, it is consistent as the number of assets tends towards infinity since $\text{plim}_{N\to\infty}\,\hat{\underline{\delta}}_t^{GLS} = \underline{\tilde{\delta}}_t$ follows when $\lim_{N\to\infty}[B'\Omega^{-1}B]^{-1} = 0$.

To understand the relationship between the generalized least squares(GLS) estimator and the minimum idiosyncratic risk estimator, it is useful to reformulate the GLS problem as a portfolio problem following Litzenberger and Ramaswamy(1979) and Rosenberg and Marathe(1979). In particular, the GLS estimator can be thought of as the solution of the following portfolio optimization procedure: choose the N portfolio weights $\underline{w}_j$ (to mimic the j^{th} factor) so that they:

$$\min_{\underline{w}_j} \underline{w}_j{}'D\underline{w}_j \tag{9a}$$

subject to:

$$\begin{aligned} \underline{w}_j'\underline{b}_k &= 0 \qquad \forall j \neq k \\ &= 1 \qquad j = k \end{aligned} \tag{9b}$$

[11] The covariance matrix of the OLS factor estimates is: $(B'B)^{-1}B'\Omega^{-1}B(B'B)^{-1}$.
[12] The covariance matrix of the GLS factor estimates is: $(B'\Omega^{-1}B)^{-1}$.

where $\underline{b}_k$ is the k^{th} column of the factor loading matrix B and D is the diagonal matrix consisting of the variances of the idiosyncratic risk vector $\tilde{\underline{\epsilon}}_t$.[13] It is straightforward to show that the solution to the portfolio problem posed in (9) is given by $(B'D^{-1}B)^{-1}B'D^{-1}$ which is equivalent to the GLS estimator given in (8) when the residual covariance matrix Ω is diagonal. In practice, we rescale the weights $\underline{w}_j$ so that they sum to one so that these sets of weights can properly be interpreted as portfolio weights. This is the generalized least squares version of the portfolio formation procedure adopted by Fama and MacBeth(1973) and similar in spirit to the GLS estimators used by Black and Scholes(1974).[14]

How does measurement error in the factor loadings and idiosyncratic variances affect the properties of the Fama-MacBeth portfolios? An examination of equations (9a) and (9b) reveals three ways that measurement error can effect the construction of the Fama-MacBeth reference portfolios: (i) the use of estimated idiosyncratic variances in the calculation of the sample residual risk of the portfolios [i.e. $\underline{w}_j{}'D\underline{w}_j$], (ii) the effect of the requirement that the portfolio weights be orthogonal to the sample loadings of the common factors not being mimicked [i.e. $\underline{w}_j'\underline{b}_k = 0 \quad \forall j \neq k$], and (iii) the repercussions of the stipulation that the Fama-MacBeth basis portfolios have sample loadings of unity [i.e. $\underline{w}_j'\underline{b}_j = 1$]. The most pernicious effect of measurement error on basis portfolio performance is likely to arise from the requirement that the sample loadings of the basis portfolios equal one. Why should this be the case? First, intuition gleaned from the statistical literature suggests that the performance of these portfolios should not be markedly degraded by the presence of measurement error in the residual variances. This sanguine conclusion follows from experience in heteroskedastic regression settings that suggests weighted least squares estimation with weights that are imperfectly correlated with the true weights *typically* achieves much of the potential gain in efficiency. Second, the requirement that the portfolio weights be orthogonal to the sample loadings of the other common factors is essentially costless since this constraint merely determines a particular *sample* rotation of the factors. Unfortunately, the stipulation that the Fama-MacBeth basis portfolios have sample loadings of unity [i.e.

[13] Note that since we are ignoring the off-diagonal elements of Ω such as industry effects, this procedure is actually better characterized as weighted least squares or diagonal generalized least squares.

[14] We will follow common usage and refer to the GLS estimator as a Fama-MacBeth estimator despite the fact that they actually used ordinary least squares.

$\underline{w}_j'\underline{b}_j = 1$] is a potential source of difficulty in the presence of measurement error in the factor loadings.[15]

The basic problem is that measurement error in the loadings will reduce the correlation of the returns on the Fama-MacBeth portfolios with the underlying common factors.[16] Why does this occur? The intuition lies in the fact that the Fama-MacBeth procedure will tend to give greater weight to security returns associated with large sample factor loadings and will typically downweight those with small loading estimates. This is appropriate in the absence of measurement error since the returns of securities with large factor loadings are more informative about fluctuations in the common factor. However, large estimated factor loadings can occur for a combination of two reasons: large true factor loadings or large errors in the estimation of small true factor loadings. Similarly, small sample factor loadings can reflect either small true factor loadings or offsetting measurement error in otherwise large loadings. Thus the Fama-MacBeth weighting procedure is less fitting when large sample loadings can arise from measurement error as well. In the extreme, measurement error in factor loadings can be so malignant as to virtually eliminate the information content of these estimates regarding fluctuations in the common factors.

The minimum idiosyncratic risk procedure employed here mitigates the harmful effects of measurement error by ignoring the differing information content of individual security returns regarding fluctuations in the factors that is implicit in the sample factor loading estimates. In particular, our procedure involves choosing the portfolio weights $\underline{w}_j$ which solve:

$$\min_{\underline{w}_j} \underline{w}_j' D \underline{w}_j \tag{10a}$$

[15] It is worth noting that the deleterious effects of measurement error on basis portfolio performance do not devolve from the analogous problems in a least squares regression setting where measurement error in the independent variables leads to biased and inconsistent estimates of regression coefficients. In the present setting, there is no particular benefit associated with an estimator possessing unbiasedness [i.e. $\mathbf{E}[\hat{\underline{\delta}}_t] = \tilde{\underline{\delta}}_t$] as opposed to an alternative estimator with $\mathbf{E}[\hat{\underline{\delta}}_t^*] = \xi + \kappa\tilde{\underline{\delta}}_t$ since the correlation of the estimator with the common factors is unaffected by affine transformations and the scale of the common factors is arbitrarily determined by the normalization that $\mathbf{E}[\tilde{\underline{\delta}}_t\tilde{\underline{\delta}}_t'] = I$. As with the orthogonality constraint, *bias* in the Fama-MacBeth portfolios taken alone only implicitly delimits a particular sample rotation of the factors.

[16] This is formally shown for the case of a single common factor in Lehmann and Modest-(1985b).

subject to:

$$\underline{w}'_j \underline{b}_k = 0 \qquad \forall j \neq k$$
$$\underline{w}'_j \underline{\iota} = 1 \tag{10b}$$

where $\underline{\iota}$ is a vector of ones and where $\underline{b}_k$ is again the k^{th} column of B.[17] These portfolios are similar to Fama-MacBeth ones in that they minimize the sample idiosyncratic variance of the basis portfolios subject to the constraint that the weights be orthogonal to the sample loadings of the factors not being mimicked [i.e. $\underline{w}'_j \underline{b}_k = 0 \quad \forall j \neq k$]. The difference between the two procedures lies in the requirement that the Fama-MacBeth portfolio have a sample loading of unity on the factor being mimicked while the minimum idiosyncratic risk portfolios must simply cost a dollar. As a consequence, the minimum idiosyncratic risk procedure ignores the information in the factor loadings: a bad decision in the absence of measurement error and a potentially good choice in its presence.[18]

In Lehmann and Modest(1985b) we scrutinized comprehensive evidence regarding the comparative merits of the two procedures. The basic conclusion reached there was that the minimum idiosyncratic risk procedure performed at least as well (and usually better than) its competitors.[19] This suggests the sampling error in our factor loadings is sufficiently serious so as to render the minimum idiosyncratic risk procedure more effective in actual practice. As a consequence, we employ this method of portfolio formation in this investigation.

Finally, a note is in order regarding the excess return portfolios that are appropriate when the riskless rate version of the **APT** is correct. For each basis portfolio formation method, we constructed minimum idiosyncratic risk portfolios using that method which had weights orthogonal to B and which cost a dollar. The details are discussed in Lehmann and Modest(1985b). As noted there, the Fama-MacBeth and minimum idiosyncratic risk

[17] This estimator can be computed as follows. Let $B = (\underline{b}_1 \underline{b}_2 \ldots \underline{b}_k)$ and suppose we are interested in mimicking the j^{th} factor. The minimum idiosyncratic risk estimator is $D^{-1}B^*[B^{*\prime}D^{-1}B^*]^{-1}\underline{e}_j$ where $B^* = (\underline{b}_1 \underline{b}_2 \ldots \underline{\iota} \ldots \underline{b}_k)$ and $\underline{\iota}$ is a vector of ones in the j^{th} column.

[18] In Lehmann and Modest(1985b) we show that the correlation of the Fama-MacBeth basis portfolios with the common factors exceeds the correlation of the minimum idiosyncratic risk portfolios with the factors in the absence of measurement error, but that the ordering may be reversed in the presence of measurement error.

[19] In addition to the two methods discussed here, we also considered two quadratic programming procedures which constrained the basis portfolios to have small weights.

procedures provide identical excess return portfolios up to a factor of proportionality. As a consequence, the remarks in this subsection are not relevant for our constructed excess return portfolios which are used under the assumption that the riskless rate version of the **APT** is true, but are relevant in analyzing the results that presume the zero beta version of the **APT** is appropriate.[20]

C. Estimation Methods

Four different methods for estimating the factor loadings and idiosyncratic variances underlying the **APT** are described in this section. Two of the methods are statistically efficient but computationally costly versions of maximum likelihood factor analysis. We also examine an instrumental variables estimator and the method of principal components. As the number of securities grows large, all four methods provide consistent estimates of the factors and, as the number of observations grows large, consistent estimates of the factor loadings and idiosyncratic variances as well.[21] However, it is obviously of greater than academic interest to know whether the comparatively inefficient methods provide performance comparable to that produced by the computationally burdensome efficient estimation methods with the data available to us. This could occur because of the large cross-sections of security returns that we employ or because of good small sample properties of the comparatively inefficient estimation methods.

The primary assumption of the **APT** is that security returns are generated by a **K** factor linear structure as given by equation (4):

$$\underline{\tilde{R}}_t = \underline{E} + B\underline{\tilde{\delta}}_t + \underline{\tilde{\epsilon}}_t \tag{4}$$

Given the structure in (4) in conjunction with the assumptions in (5) and (6) about the covariance matrices of the residual risks and the factors, the covariance matrix of security returns, Σ, can be written as:

$$\Sigma = BB' + \Omega \tag{11}$$

Theoretically, the **APT** places no restrictions on Ω other than the requirement that the

---

[20] If we employed measured riskless rates (i.e. one month Treasury bill returns) instead of the excess return portfolios, the differences between the procedures would again be of considerable concern. In this circumstance, we would employ the minimum idiosyncratic risk procedure.

[21] Given sufficient assumptions on Ω.

off-diagonal elements are sufficiently sparse so that the residual risks are diversifiable (in the limit) and, hence, security returns satisfy an approximate factor structure.[22] Unfortunately, it is not possible to estimate the factor loadings and the elements of Ω when security returns possess only an approximate factor structure. One popular way to proceed is to assume that security returns satisfy an exact statistical factor structure in that residual risks are uncorrelated across firms. With this additional assumption, the residual covariance matrix Ω is equal to a diagonal matrix D, and one can proceed with estimation using the fact that the covariance matrix of security returns Σ can be written as:

$$\Sigma = BB' + D \tag{12}$$

Obviously, efficient estimation requires *a priori* specification of the joint distribution of security returns and the factors. Under the assumption of joint normality, the sample covariance matrix of security returns is distributed as Wishart, and the log likelihood function of Σ conditional on the sample covariance matrix is given by:

$$\begin{aligned}\mathcal{L}(\Sigma|S) &= \frac{-NT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(\tilde{\underline{R}}_t - \overline{\underline{R}})'\Sigma^{-1}(\tilde{\underline{R}}_t - \overline{\underline{R}}) \\ &= \frac{-NT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{T}{2}\mathbf{trace}(S\Sigma^{-1})\end{aligned} \tag{13}$$

Maximization of the log likelihood function (13) subject to the covariance restriction given by (12), provides efficient estimates of the factor loadings and idiosyncratic variances underlying the presumed statistical factor analysis model of security returns. Under the null hypothesis that the **APT** is true, however, there is additional information in the theoretical restriction given in equation (3) that expected security returns are spanned by their factor loadings and the factor risk premia which can, in principle, lead to more efficient estimates of B and D. Imposition of this additional constraint involves taking the log likelihood function (13) and substituting in the **APT** mean restriction:

$$\underline{E} = \underline{\iota}\lambda_0 + B\underline{\lambda} \tag{14}$$

[22] The formal requirement is that as $N \to \infty$ the eigenvalues of Ω remain bounded.

The log likelihood function is then given by:

$$\mathcal{L}(\Sigma|S) = \frac{-NT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(\tilde{\underline{R}}_t - [\underline{\iota}\lambda_0 + B\underline{\lambda}])'\Sigma^{-1}(\tilde{\underline{R}}_t - [\underline{\iota}\lambda_0 + B\underline{\lambda}]) \tag{15}$$

which can be rewritten as:

$$\mathcal{L}(\Sigma|S) = \frac{-NT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}[(\tilde{\underline{R}}_t - \overline{\underline{R}}) + (\overline{\underline{R}} - [\underline{\iota}\lambda_0 + B\underline{\lambda}])]'\Sigma^{-1}[(\tilde{\underline{R}}_t - \overline{\underline{R}}) + (\overline{\underline{R}} - [\underline{\iota}\lambda_0 + B\underline{\lambda}])] \tag{16}$$

This can be simplified to:

$$\mathcal{L}(\Sigma|S) = \frac{-NT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{T}{2}\mathbf{trace}(S\Sigma^{-1}) - \frac{T}{2}(\overline{\underline{R}} - \underline{\iota}\lambda_0 - B\underline{\lambda})'\Sigma^{-1}(\overline{\underline{R}} - \underline{\iota}\lambda_0 - B\underline{\lambda}) \tag{17}$$

Maximization of (17) yields what we refer to below as restricted maximum likelihood estimates. This restricted maximum likelihood procedure involves maximizing the unrestricted log likelihood function (13) plus an additional term involving the weighted average of the deviations of the sample mean security returns from the product of the factor loadings and the corresponding risk premia.

The unrestricted maximum likelihood estimates of the relevant parameters may be obtained by setting the derivatives of (13) equal to zero and iteratively solving the first order conditions with respect to B and D. The corresponding restricted maximum likelihood estimates may be obtained by setting the derivatives of (17) equal to zero and iteratively solving the first order conditions with respect to $B, D, \underline{\lambda}$, and λ_0. While this is a conceptually simple exercise, it is computationally infeasible to obtain these estimates by iteratively solving the first order conditions when the number of securities being analyzed is substantial. We therefore employed a significantly cheaper alternative: the **EM** algorithm of Dempster, Laird, and Rubin(1977). In Lehmann and Modest(1985a) we discuss the actual estimation procedure. Its primary virtue is that its memory requirements and computational costs are less burdensome than the standard procedures for performing maximum likelihood analysis, such as those described in Lawley and Maxwell(1971).

The maximum likelihood procedures described above provide efficient estimates of the factor loadings and idiosyncratic variances used as inputs in (10) to construct portfolios that mimic the common factors. The principal drawback of these procedures is that they are moderately expensive in terms of the real computer time it takes for each run to be completed. Given the relatively high cost of these efficient procedures, several authors have proposed the use of less costly procedures with the hope that the loss in efficiency is only small. Chamberlain and Rothschild(1983) and Connor and Korajczyk(1984), for instance, have recently suggested the use of principal components as an inexpensive alternative to maximum likelihood factor analysis. The connection between principal components and the statistical factor analysis model is, as Chamberlain and Rothschild(1983) showed, that as the number of securities being analyzed tends toward infinity, the first K eigenvectors obtained from the eigenvalue decomposition of the covariance matrix of security returns will converge to the factor loadings underlying security returns. Connor and Korajczyk(1984) showed that this holds for the sample covariance matrix as well. The primary disadvantage of principal components is that it ignores information contained in the sample idiosyncratic variances which potentially might lead to more efficient estimates of the factor loadings in a finite sample.[23] Nonetheless, since the one time extraction of eigenvalues and eigenvectors does not require iteration over first order conditions like the nonlinear maximum likelihood procedures, it can be cheaper and therefore a potentially attractive alternative.[24]

We employed the singular value decomposition algorithm included in the NAG Subroutine Library to obtain the required eigenvalues and eigenvectors. Each column of the **K** eigenvectors was multiplied by the square root of the corresponding eigenvalue in order to scale the factors to have unit variance. Estimates of the idiosyncratic variances were then obtained using equation (12) to solve for D given the the estimated factor loadings B (the

[23] This can be seen by noting that principal components is equivalent to maximum likelihood factor analysis when the idiosyncratic variances are assumed to be identical, i.e. when: $D = \sigma^2 I$. Thus the relationship between principal components and maximum likelihood is similar to the relationship between ordinary least squares and generalized least squares.

[24] Principal components is cheaper in that one eigenvalue decomposition provides the eigenvectors (and hence the factor loadings) needed for a factor model with anywhere between 1 and N factors whereas this would require N different maximum likelihood runs. However, unrestricted maximum likelihood factor analysis of 750 securities with 5 factors using the **EM** algorithm, for instance, would on average be less costly than principal components although significantly more expensive than one instrumental variables run with 5 factors on 750 securities.

transformed eigenvectors), and the sample covariance matrix S.

Another relatively inexpensive alternative to maximum likelihood factor analysis is instrumental variables factor analysis along the lines suggested by Madansky(1964) and Hagglund(1982). Instrumental variables estimators have recently been employed by Chen(1983) and Madansky and Marsh(1985). The central idea of these instrumental variables estimators is to substitute consistent estimates of the factors $\tilde{\underline{\delta}}_t$ for the factors themselves in equation (4) and then estimate the factor loadings B by ordinary least squares.[25] Chen(1983) used portfolios formed by mathematical programming based on maximum likelihood factor analysis of 180 securities as the required consistent estimates of the factors. We employ a simpler instrumental variables procedure, which is described in detail in Lehmann and Modest(1985b) that does not require preliminary maximum likelihood factor analysis.

III. On The Detection of Abnormal Performance

The ability to construct benchmark portfolios along the lines of (10) which are potentially free of the biases which have been attributed to the **CAPM** benchmarks suggests their use in performance evaluation. However, the putative freedom from benchmark error of the reference portfolios need not imply that their use in the usual strategies for assessing mutual fund performance will lead to correct inferences. As is obvious from the discussion in Mayers and Rice(1979), Verrechia(1980), Dybvig and Ross(1981), Admati and Ross(1985), Pfleiderer and Bhattacharya(1983), and Grinblatt and Titman(1985), benchmark error is not the only difficulty plaguing performance evaluation. Nontrivial problems of inference arise when mutual fund managers have some ability to predict benchmark returns.

If a portfolio manager has substantive market timing ability, the manager's portfolio choices will be correlated with the subsequent benchmark returns. Hence, measured covariation between the mutual fund returns and benchmark returns will convolve two influences — the manager's market timing ability and the chosen level of risk. Only in special circumstances will it be possible to sort out these effects. Since measured covariation need not

[25] The application of instrumental variables methods to factor analysis models typically involves the assumption that the idiosyncratic disturbances are independent (as in the statistical factor analysis model). The procedures, however, will provide consistent estimates even when the idiosyncratic disturbances are correlated so long as the disturbances are sufficiently independent for a law of large numbers to apply as the number of securities tends toward infinity.

only reflect the *(constant)* risk level chosen by the manager, it cannot be used to naively form a combination of reference portfolios which are of comparable risk to the mutual fund. Thus it will not be possible to rate managerial investment performance with these tools in the presence of market timing ability.

The problem is considerably simplified if our goal is merely the detection of abnormal performance and not the measurement of its degree. Excess returns regressions of the form employed by Jensen(1968,1969) will, apart from sampling error, successfully detect abnormal performance even if managers possess market timing ability. Hence we will limit our attention to this more modest goal and defer for the present the quantification of managers investment skills.

Suppose that the uninformed investors perceive that the **APT** is exactly true so that returns on N individual securities satisfy:

$$\tilde{\underline{R}}_t = B\tilde{\underline{R}}_{mt} + \tilde{\underline{\epsilon}}_t \tag{18}$$

where $\tilde{\underline{R}}_{mt}$ is a $K \times 1$ vector of reference portfolios and B is the $N \times K$ matrix of factor loadings. Here we let $\tilde{\underline{R}}_t$ and $\tilde{\underline{R}}_{mt}$ denote raw returns when the zero beta version of the **APT** is appropriate and we presume that they represent excess returns when the riskless rate formulation is correct. We assume that the reference portfolios $\tilde{\underline{R}}_{mt}$ are perfectly correlated with the common factors and that B is measured without error as well. Consider the return on a mutual fund portfolio:

$$\begin{aligned} \tilde{R}_{pt} &= \sum_{i=1}^{N} \omega_i(\underline{s}_t)\tilde{R}_{it} \\ &= \sum_{i=1}^{N} [\omega_i(\underline{s}_t)\underline{b}_i'\tilde{\underline{R}}_{mt} + \omega_i(\underline{s}_t)\tilde{\epsilon}_{it}] \end{aligned} \tag{19}$$

where $\omega_i(\underline{s}_t)$ is the weight of security i in the mutual fund portfolio at date t, $\underline{b}_i'$ is a $1 \times K$ row vector of B and $\underline{s}_t$ is a vector of signals received by the mutual fund manager that are used for predicting $\tilde{\underline{R}}_{mt}$ and $\tilde{\underline{\epsilon}}_t$. It is convenient to rewrite (19) as:

$$\tilde{R}_{pt} = \underline{\beta}'_{pt}\tilde{\underline{R}}_{mt} + \tilde{\epsilon}_{pt} \tag{20}$$

where:

$$
\begin{aligned}
\underline{\beta}'_{pt} &= \underline{\bar{\beta}}'_p + \underline{x}(\underline{s}_t)' \\
&= \sum_{i=1}^{N} \omega_i(\underline{s}_t)\underline{b}'_i \\
\tilde{\epsilon}_{pt} &= \sum_{i=1}^{N} \omega_i(\underline{s}_t)\tilde{\epsilon}_{it}
\end{aligned}
\tag{21}
$$

In (21), $\underline{\bar{\beta}}_p$ is the target or average $\underline{\beta}$ of the fund, and $\underline{x}(\underline{s}_t)$ is the time t deviation from $\underline{\bar{\beta}}_p$ selected by the manager (assumed to average zero over the sample). Note that if the manager possesses stock selection ability, $\tilde{\epsilon}_{pt}$ will not have a zero mean.

What happens if, as uninformed investors, we run the regression of $\tilde{R}_{pt}$ on $\underline{\tilde{R}}_{mt}$? Letting $\mathbf{E}^*[X|Y]$ denote the minimum variance linear estimator of X given Y (i.e. the regression function), we obtain:

$$
\mathbf{E}^*[\tilde{R}_{pt}|\underline{\tilde{R}}_{mt}] = \hat{\alpha}_p + \underline{\hat{\beta}}'_p \underline{\tilde{R}}_{mt}
\tag{22}
$$

where:

$$
\begin{aligned}
\hat{\alpha}_p &= [\bar{\epsilon}_p - \mathrm{Cov}\{\underline{x}'_t\underline{\tilde{R}}_{mt}, \underline{\tilde{R}}'_{mt}\}'\Sigma_m^{-1}\underline{\bar{R}}_m + \mathbf{E}\{\underline{x}'_t\underline{\tilde{R}}_{mt}\}] \\
\underline{\hat{\beta}}_p &= [\underline{\bar{\beta}}_p + \Sigma_m^{-1}\mathrm{Cov}\{\underline{x}'_t\underline{\tilde{R}}_{mt}, \underline{\tilde{R}}_{mt}\}] \\
\bar{\epsilon}_p &= \sum_{i=1}^{N} \mathrm{Cov}\{\omega_i(\underline{s}_t), \tilde{\epsilon}_{it}\} \\
\Sigma_m &= \mathbf{E}[\{\underline{\tilde{R}}_{mt} - \underline{\bar{R}}_m\}\{\underline{\tilde{R}}_{mt} - \underline{\bar{R}}_m\}'] \\
\underline{\bar{R}}_m &= \mathbf{E}[\underline{\tilde{R}}_{mt}]
\end{aligned}
\tag{23}
$$

and $\underline{x}'_t$ is used as shorthand notation for $\underline{x}(\underline{s}_t)'$ and $\mathrm{Cov}\{\underline{x}'_t\underline{\tilde{R}}_{mt}, \underline{\tilde{R}}_{mt}\}$ is a $1 \times K$ vector of the covariances between $\underline{x}'_t\underline{\tilde{R}}_{mt}$ and the K elements of $\underline{\tilde{R}}_{mt}$. The coefficient $\hat{\alpha}_p$ is the usual Jensen performance measure.

In the absence of the ability to pick stocks [i.e. $\bar{\epsilon}_p = 0$] and to time the market [i.e. $\mathbf{E}\{\underline{x}'_t\underline{\tilde{R}}_{mt}\} = \mathrm{Cov}\{\underline{x}'_t\underline{\tilde{R}}_{mt}, \tilde{R}_{jt}\} = 0$ for all j=1,...K], the regression equation (22) will indicate no abnormal performance since, in this instance:

$$
\mathbf{E}^*[\tilde{R}_{pt}|\underline{\tilde{R}}_{mt}] = \underline{\bar{\beta}}'_p \underline{\tilde{R}}_{mt}
\tag{24}
$$

If the mutual fund manager possesses stock selection ability but no market timing ability, the regression will indicate superior performance since :

$$\mathbf{E}^{\star}[\tilde{R}_{pt}|\underline{\tilde{R}}_{mt}] = \bar{\epsilon}_p + \underline{\bar{\beta}}'_p \underline{\tilde{R}}_{mt} \tag{25}$$

where $\bar{\epsilon}_p > 0$ under mild restrictions. This is merely a restatement of the Mayers and Rice (1979) proposition, as simplified and extended by Dybvig and Ross(1981), that the Jensen measure will correctly indicate superior performance when managers possess security selection ability but are unable to time the market. Finally, if portfolio managers possess market timing ability as well, the Jensen measure may be positive or negative depending on the terms in brackets on the first line of (23). Hence, the Jensen measure will indicate abnormal performance but cannot be used to rank managers.

These results are useful in that the simple Jensen measure can, ignoring sampling variation, detect both normal and abnormal performance. Unfortunately, it is not capable of indicating whether managerial ability is of the market timing or stock selection variety. Yet there is a hint in (22) and (23) of the possibility of detecting the presence of market timing ability due to the terms involving $\mathrm{Cov}\{\underline{x}'_t\underline{\tilde{R}}_{mt}, \underline{\tilde{R}}_{mt}\}$ and $\mathbf{E}\{\underline{x}'_t\underline{\tilde{R}}_{mt}\}$. These terms suggest that perhaps a quadratic regression could detect market timing ability when returns unconditionally follow the **APT** (18).

The quadratic regression framework originally was examined by Treynor and Mazuy (1966). Its possibilities as a framework for separating market timing and stock selection ability were studied by Jensen(1972), an analysis which was corrected and extended in Pfleiderer and Bhattacharya(1983). Its possibilities are seen by considering (for the sake of notational simplicity) the one factor version of (20):

$$\tilde{R}_{pt} = \beta_{pt}\tilde{R}_{mt} + \tilde{\epsilon}_{pt} \tag{26}$$

and studying the quadratic regression:

$$\mathbf{E}^{\star}[\tilde{R}_{pt}|\tilde{R}_{mt}, \tilde{R}^2_{mt}] = \alpha^{\star}_p + b^{\star}_{1p}\tilde{R}_{mt} + b^{\star}_{2p}\tilde{R}^2_{mt} \tag{27}$$

The regression slope coefficients are given by:

$$
\begin{aligned}
\begin{bmatrix} b^*_{1p} \\ b^*_{2p} \end{bmatrix} &= \left(\text{Var} \left\{ \begin{matrix} \tilde{R}_{mt} \\ \tilde{R}^2_{mt} \end{matrix} \right\} \right)^{-1} \text{Cov} \left[\tilde{R}_{pt}, \begin{pmatrix} \tilde{R}_{mt} \\ \tilde{R}^2_{mt} \end{pmatrix} \right] \\
&= \begin{pmatrix} \sigma^2_m & \sigma_{3m} \\ \sigma_{3m} & \sigma_{4m} \end{pmatrix}^{-1} \begin{bmatrix} \bar{\beta}_p \sigma^2_m + \text{Cov}(\tilde{x}_t, \tilde{R}^2_{mt}) \\ \bar{\beta}_p \sigma_{3m} + \text{Cov}(\tilde{x}_t, \tilde{R}^3_{mt}) \end{bmatrix} \\
&= \begin{bmatrix} \bar{\beta}_p \\ 0 \end{bmatrix} + \frac{1}{\sigma^2_m \sigma_{4m} - \sigma^2_{3m}} \begin{bmatrix} \sigma_{4m} & -\sigma_{3m} \\ -\sigma_{3m} & \sigma^2_m \end{bmatrix} \begin{bmatrix} \text{Cov}(\tilde{x}_t, \tilde{R}^2_{mt}) \\ \text{Cov}(\tilde{x}_t, \tilde{R}^3_{mt}) \end{bmatrix} \\
&\equiv \begin{bmatrix} \bar{\beta}_p \\ 0 \end{bmatrix} + \begin{bmatrix} \gamma_{1p} \\ \gamma_{2p} \end{bmatrix}
\end{aligned} \tag{28}
$$

where σ_{3m} and σ_{4m} are the skewness and kurtosis of R_{mt}, respectively, and $\bar{\beta}_p$ is the target β of the mutual fund. Similarly, the intercept of the quadratic regression is:

$$
\begin{aligned}
\alpha^*_p &= \bar{\epsilon}_p + \bar{\beta}_p \bar{R}_m + \text{Cov}(\tilde{x}_t, \tilde{R}_{mt}) - b^*_{1p} \bar{R}_m - b^*_{2p} \bar{R}^2_m \\
&= \bar{\epsilon}_p + \text{Cov}(\tilde{x}_t, \tilde{R}_{mt}) - \gamma_{1p} \bar{R}_m - \gamma_{2p} \bar{R}^2_m
\end{aligned} \tag{29}
$$

In the absence of market timing ability, $\text{Cov}\{x_t, R^2_{mt}\}$ and $\text{Cov}\{x_t, R^3_{mt}\}$ are both zero so that the coefficient on R_{mt} will be the target beta of the fund and that on R^2_{mt} will be zero. So long as the market timing information and preferences of the manager and the distribution of R_{mt} are such that the appropriate combination of $\text{Cov}\{x_t, R^2_{mt}\}$, $\text{Cov}\{x_t, R^3_{mt}\}$, σ_{3m}, and σ_{4m} are non-zero, then b^*_{2p} will be non-zero, indicating the presence of market timing ability.

This is a remarkable result in that it makes no assumptions about the return generating process beyond stationarity, linearity and the validity of the **APT** benchmark. Without assuming anything beyond finite third and fourth order moments, the quadratic term b^*_{2p} will be zero in the absence of timing ability. So long as the information and the preferences of the fund manager are such that portfolio returns are skewed to the right, as conceived of, for example, in the first example of Dybvig and Ross(1981), b^*_{2p} will be non-zero and, hence, will indicate the presence of timing ability. The basis idea is quite simple: market timers should make money when the market rises or falls dramatically, that is, when the squared return on the market is large.

Of course, without further restriction on distributions and preferences, it will not, in general, be possible to measure the magnitudes of market timing and security selection ability. As is obvious from equation (28) above, if there is no co-skewness between the fluctuations in the fund beta and the return on the factor (i.e. $\mathrm{Cov}\{x_t, R_{mt}^2\}$ is zero), it will be possible to estimate the target beta of the fund and $\mathrm{Cov}\{x_t, R_{mt}^3\}$, but it will not be possible to separate the two sources of abnormal performance. This can be accomplished by placing sufficient structure on the problem so that measurement of $\mathrm{Cov}\{x_t, R_{mt}^3\}$ leads to estimation of $\mathrm{Cov}\{x_t, R_{mt}\}$, which permits the estimation of $\bar{\epsilon}_p$ using equation (29). For example, Pfleiderer and Bhattacharya(1983) assume the joint normality of $\underline{\tilde{R}}_{mt}, \underline{s}_t$, and $\tilde{\epsilon}_{pt}$ and the linearity of $\underline{x}_t$ in $\underline{s}_t$. Still, the potential capability to detect the presence of market timing ability with the simple quadratic regression procedures represents a promising advance, although its actual usefulness is, of course, an empirical question that we will begin to examine below.

IV. The Data

In implementing the **CAPM** to obtain risk-adjusted excess returns, we use the standard benchmarks: the CRSP equally-weighted and value-weighted indices of NYSE stocks taken from the CRSP monthly index file. The construction of reference portfolios for the **APT** benchmarks, however, is not so straightforward. Some of the computational aspects and potential virtues of efficient estimation procedures were detailed in the preceding sections. Here we discuss some of the data analytic alternatives facing investigators.

One choice facing researchers is the appropriate frequency of observation for estimating the factor models of security returns underlying the **APT**. Many choices are readily available since the **CRSP** daily return file provides returns on all **NYSE** and **AMEX** stocks since July 1962, and minimal computational skill stands between us and weekly, monthly or other intermediate frequency data. The primary advantage of daily data is the potential increase in precision of the estimated variances and covariances, the inputs to the factor analysis model, that comes with sampling the data more often. The two main disadvantages of daily data are the persistent incidence of non-trading and thin trading which bias the estimates of second-order moments and the biases in mean returns associated with bid-ask spreads that are studied in Blume and Stambaugh(1983) and Roll(1983). Following Roll and Ross(1980), we opted for the putative benefits of a large sample and used daily data to estimate the factor models, although the optimal observation frequency is an empirical question worthy of detailed investigation and one which we are currently examining (Lehmann and Modest(1985c)). Portfolio weights constructed from daily data based on the minimum idiosyncratic risk procedure outlined above were then multiplied by monthly security returns to construct monthly returns on our basis portfolios.[26]

Our mutual fund data base consists of the returns on one hundred and thirty (130) mutual funds over the fifteen year period January 1968-December 1982. We are grateful to Roy Henriksson for graciously supplying us with the vast majority of this data. The monthly returns are calculated from the end of month bid prices and monthly dividends obtained from Standard and Poor's *Over-the-Counter Daily Stock Price Record*, Weisenberger's *Investment Companies* annual compendium, and Moody's *Annual Dividend Record*. It is worth emphasizing that the *Over-the-Counter Daily Stock Price Record* omits a significant fraction of the dividends paid and reliance must be made on the other two sources to obtain accurate dividend information. The sample was chosen to include a variety of funds with differing risk postures. No municipal bond fund or option fund, however, was included. It should also be pointed out that the Henriksson sample was chosen so that the funds survived over the January 1968-June 1980 period. This raises the potential for some problems due to a possible survivorship bias, although the results below show no evidence of such a bias. Due to our concern that the risk levels of the funds were not constant over the fifteen year period, we restricted our attention to examining the behavior of the funds over three five-year subperiods: January 1968-December 1972, January 1973-December 1977, and January 1978-December 1982.

The **CRSP** daily file contains 1359, 1346 and 1281 securities which were continuously listed and had no missing observations during the three five-year periods covered by our mutual fund data. We confined our attention to these firms in order to have the same number of observations for each security and ignored any potential selection bias associated with this choice. Computational considerations required the analysis of no more than 750 securities simultaneously. We have carried out runs using as many as one thousand

[26] Another alternative would have been to multiply the portfolio weights times daily security returns and then aggregate these daily returns to obtain monthly portfolio returns. In Lehmann and Modest(1985b), we examined both reference portfolio formation procedures and found little difference between the two approaches. Concern over bid-ask spread bias in this procedure, which calls for daily rebalancing, led us to opt for the buy-and-hold strategy.

securities. However, the larger number of securities yielded a minimal improvement over the performance of reference portfolios based on 750 securities and proved to be disproportionately expensive in terms of the computational time. The **CRSP** daily file lists securities in alphabetical order by most recent name. To guard against any biases induced by the natural progression of letters (General Electric, General Motors, etc.), we randomly reordered the firms. The number of daily observations in these samples was 1234, 1263, and 1264 respectively. The usual sample covariance matrix of these security returns provided the basic input to our subsequent analysis.

V. Empirical Results

In this section we provide evidence on the comparative performance of different benchmarks for evaluating mutual fund performance. In particular, do the absolute and relative rankings of the funds hinge on which benchmark is chosen to evaluate normal performance? Tables 1 – 6 provide evidence on the performance of alternative **APT** benchmarks. The tables summarize and contrast the behavior of the intercepts from simple Jensen-style regressions of mutual fund returns on the **APT** basis portfolios as given by equation (22). The basic questions here are whether the less efficient (and less costly) basis portfolio formation procedures lead to different conclusions than the more efficient procedures and whether the conclusions about performance are sensitive to the number of factors assumed to underly security returns or the number of securities included in the analysis. Tables 7 and 8 provide the corresponding information comparing **APT** and **CAPM** benchmarks in order to highlight the contrasts across asset pricing models. Table 9 summarizes the information from quadratic regressions along the lines of equation (27) using both **APT** and **CAPM** benchmarks in order to shed some light on one possible cause of the anomalous behavior of the intercepts from the Jensen-style mutual fund regressions.[27]

The first eight tables provide careful scrutiny of the similarities and differences in the intercepts across performance benchmarks. The tables come in pairs (i.e. the eight tables consist of four sets of two tables each), each of which contrasts the behavior of different benchmarks in a particular dimension. The first of the two tables summarize the central

[27] In the current version of the paper we only include the results from the quadratic regressions using the five factor **APT** benchmark.

tendencies of the intercepts from each benchmark across the 130 mutual funds. The second table in each pair provides two measures describing the relationships among the intercepts from the different benchmarks.

The first table in each pair provides four summary measures describing the typical behavior of the intercepts for each benchmark over our three sample periods. The first three statistics are the mean intercept, the mean absolute intercept, and the average absolute t-statistic.[28] These t-statistics are simply the estimated intercepts divided by the usual ordinary least squares standard errors where the standard errors are calculated under the assumption that the residuals in (22) are independent and have common variance over time. Unfortunately, these t-statistics may lack the appropriate statistical justification in the context of mutual fund performance evaluation: as long as managers vary the composition of their portfolios in attempts to outperform the market, mutual fund returns will be likely to have non-stationary variances even if the return generating process of individual securities is stationary and, hence, the resulting heteroskedasticity biases the associated t-statistics.[29] To guard against this possibility, we also present adjusted t-statistics using estimated standard errors that are consistent in the presence of arbitrary forms of heteroskedasticity, rather than the usual least squares standard errors. These adjusted standard errors have been proposed and examined in work by Hansen(1982), White(1980) and Hsieh(1983).[30]

Of course, these usual summary statistics merely serve to characterize the typical behavior of these *sample* intercepts and cannot be used to draw inferences about the typical behavior of the *true* intercepts without further assumptions that facilitate statistical inference. Unfortunately, we are unable to construct the usual joint *F-test* of whether the intercepts are significantly different from zero since the number of funds (130) is greater than the number of time series degrees of freedom (60) and hence the sample covariance matrix of the residuals from equation (20) is singular. One strategy to employ in the face of this difficulty is to examine the individual t-statistics to see if any of them are sufficiently

---

[28] In parentheses below the sample averages we also present the sample standard deviations of the intercepts and the t-statistics across funds.

[29] In a slightly different context, the problems associated with heteroskedasticity in the evaluation of mutual fund performance are discussed in more detail by Breen, Jagannathan, and Ofer(1984).

[30] An analysis of their small sample properties using Monte Carlo simulations is presented in Hsieh(1983).

large that it would surely lead to an F-statistic greater than the appropriate critical value. This is the intuition behind the application of the Bonferroni inequality to this problem as discussed by Miller(1966) and employed in Fama(1984). Briefly, the Bonferroni inequality states that if we examine N possibly dependent t-statistics at the critical value associated with α/N, then we are sure that we have at most a joint test at the significance level α. Of course, the inherent conservatism in reducing the significance level to α/N means that we will often fail to reject the null hypothesis when it is false. This does not appear to be a problem in this application, however, as in each period there are a number of t-values greater than four which is high enough to violate the Bonferroni bound at the one per cent level.

The second table in each pair describes the degree of association among the intercepts computed from different performance benchmarks.[31] We employ two measures to characterize these relationships: simple (i.e. Pearson product moment) correlations and Spearman rank correlations.[32] The simple correlations are well-known statistics which have a variety of conventional interpretations. For example, the squared simple correlation between intercepts from two different benchmarks is a (biased) estimate of the percentage of the variation in the intercepts from one benchmark that can be explained by variation in the intercepts from a second procedure **(i.e.** $[1 - \frac{\textbf{unexplained variation}}{\textbf{total variation}}]$ **from a simple cross-sectional regression of the intercepts from one benchmark on those of another)**. One difficulty in interpreting simple correlations is that they are very sensitive to outlying observations—a small number of large positively related outliers may cause a

[31] It is argued by some that ranking funds on the basis of estimated alphas is inappropriate since differences may merely reflect, for instance, diversity in leverage and that a more appropriate measure is the Treynor-Black appraisal ratio: the estimated intercept divided by the idiosyncratic variation. Since the goal here is simply to see whether rankings depend on which benchmark is chosen, the question of which ranking scheme is more appropriate is irrelevant. To the extent that the conventional wisdom is correct and alternative risk adjustment procedures lead to similar inferences, the rankings should be the same across benchmarks regardless of the particular ranking scheme chosen. Since the intercepts are economic quantities which are of considerable interest in their own right, we provide the evidence in this form. Moreover, we suspect that we would obtain similar results if we worked with the Treynor-Black appraisal ratio or, equivalently, the t-statistics of the intercepts.

[32] Unfortunately, we cannot report standard errors and confidence intervals for our estimated rank and simple correlations due to the likely presence of correlation among the intercepts across funds. As a consequence, the correlations cannot be subjected to formal statistical tests without further assumptions.

sample correlation to be large and positive even when there is a strong negative relationship between the bulk of the observations. This can occur because, in the presence of outliers, the total variation in intercepts across funds is likely to be relatively large compared to the unexplained variation in intercepts since the bulk of the total variation will be explained by the similar behavior of the outliers. Since the intercepts from different benchmarks typically are of similar magnitude for the funds with very large positive and negative alphas, the sensitivity of simple correlations to extreme observations is a potentially serious problem. Consequently, we also report Spearman rank correlations, which provide estimates of the degree of association among intercepts which have little sensitivity to outlying observations.

Not surprisingly, rank correlations measure the degree of association of the ranks of the intercepts (i.e. the firm with the largest alpha is ranked one, that with the second largest alpha is ranked two, etc.) across the different benchmarks. Their insensitivity to outlying observations follows from measuring the magnitude of an intercept by its rank: the difference in the size of the largest alpha and the tenth largest alpha can be enormous, but the difference in ranks is only nine. The use of rank correlations has another justification in this context. Since one purpose of mutual fund performance evaluation is to provide ordinal rankings of funds, rank correlations summarize differences in the inferences produced by alternative benchmarks in an apt way.

Since rank correlations are less widely employed among financial economists than simple correlations, one natural question concerns the interpretation of the magnitudes of sample rank correlations. One answer to this question in the present setting can be obtained by examining the formula for the Spearman rank correlation:

$$\begin{aligned}\rho_{\mathrm{jk}}^{\mathrm{rank}} &= 1 - \frac{6\sum_{i=1}^{N}(y_{ij} - y_{ik})^2}{(N(N^2-1))} \\ &= 1 - \frac{6 * \text{Average Squared Difference of Ranks}}{(N^2-1)}\end{aligned} \tag{30}$$

where y_{ij} is the rank of the i^{th} firm using the j^{th} benchmark. Obviously, the sample variance of the difference in ranks[33] of the intercepts is the key variable (the remaining numbers simply transform the sample variance of the difference in ranks into an approximately

33 The average squared deviation of the sample difference in ranks equals the sample variance since the average difference in ranks must equal zero.

unbiased estimate of the simple correlation). With 130 funds, if the typical difference in ranks is one (i.e. funds with ranks 1 and 2 by one method are ranked 2 and 1 by the other method, funds with ranks 3 and 4 switch places, etc.), the sample variance of the difference in ranks will be one as well and the rank correlation would equal 0.9996. Similarly, if the typical difference in the ranks of intercepts between two methods was two and the corresponding sample variance of these differences is four, the rank correlation would equal 0.9986. Other rank correlations associated with different typical rank differences are easy to compute—a typical rank difference of five ranks leads to a rank correlation of 0.9911, a ten rank difference implies a 0.9645 rank correlation, fifteen implies 0.9201, twenty implies 0.8580, twenty-five implies 0.7781, thirty implies 0.6805, thirty-five implies 0.5651, forty implies 0.4319, forty-five implies 0.2810, and fifty implies 0.1124. Clearly, rank correlations of 0.5 to 0.8 are associated with very large typical deviations in the ranks of intercepts across benchmarks.

The first thing one notices in the Jensen-type regressions reported in the Tables is the persistent incidence of negative intercepts especially with the **APT** benchmarks. Consider, for instance, the average estimated alphas from the regressions run in excess return form using the unrestricted maximum likelihood estimation procedure presented in Table 1 (which we simply refer to as maximum likelihood in the tables). The average excess return of the funds was −4.76% on an annual basis for the first five year period (−.406% per month), −6.14% annually for the second period (−.527% per month), and −1.50% per year for the third period(−.126% per month).[34] Inspection of the individual alphas of the funds verifies that the intercepts are almost uniformly negative and that the mean is not being pulled down by a few funds with exceptionally poor performance. This can also be seen by noting the relatively small difference between the absolute values of the average intercepts and the sample means of the absolute values of the intercepts. In the first two five year periods the average values of the adjusted and unadjusted t-statistics are sufficiently large to suggest that many of the intercepts are significantly different from zero. In short, our

[34] Since the conclusions one would reach about the performance of alternative **APT** benchmarks turns out to be independent of whether the regressions are run in raw or excess return form, we limit ou discussion in the text to the results from regressions run in excess return form. We emphasize these results due to evidence presented in Lehmann and Modest(1985a) that suggests the preferability of this form of the **APT**.

APT benchmarks suggest the presence of widespread abnormal performance by mutual funds across our sample periods, a finding which we further discuss below.

With these preliminary observations in hand, Tables 1 and 2 examine the impact of alternative methods of estimating the factor model for security returns underlying the **APT**. Four different estimation methods were compared using samples of 750 securities: two maximum likelihood procedures, an instrumental variables estimator and the method of principal components. These procedures were discussed in Section II.B. The maximum likelihood procedures differ in that the restricted maximum likelihood method also uses the information in the sample means as outlined in equation ($15a$) while conventional maximum likelihood factor analysis ignores this information. Tables 1 and 2 compare these **APT** benchmarks assuming there are five common sources of systematic risk.[35] For each estimation method, statistics are presented for regressions run in raw return form, which corresponds to the zero-beta version of the **APT**, and for regressions run in excess return form corresponding to the riskless rate version of the **APT**.

Examination of Table 1 reveals that there is some variation in the mean intercept across estimation methods, although the differences in most cases are not terribly large. Compared with the average intercept using the unrestricted maximum likelihood procedure, the mean alpha from the restricted maximum likelihood procedure was 28 basis points per annum lower in the first five year period, 2 basis points higher in the second period and 5 basis points per year higher in the third period. Again using the average intercepts from the unrestricted maximum likelihood procedure as a standard of comparison, the average intercepts using the instrumental variables estimation procedure were 35 basis points higher, 16 basis points lower, and 63 basis points higher in periods one through three respectively. The corresponding numbers using the principal components procedure were +208 basis points, +29 basis points, and +11 basis points.

Table 2 presents additional information on the absolute and relative rankings across estimation methods in the form of simple and Spearman rank correlation coefficients between the intercepts using the different estimation methods. As is evident from the Table, the intercepts from the unrestricted and restricted maximum likelihood procedures are al-

[35] Tables A1 and A2 in the Appendix present the corresponding results when there are assumed to be ten common sources of covariation.

ways highly correlated. The lowest simple correlation between the intercepts in the three periods is .9993, which occurs in the first five year period. For all three periods, the rank correlation is .9999. The degree of similarity between the absolute and relative rankings using the unrestricted maximum likelihood procedure and the less efficient instrumental variables and principal components procedures is not nearly so strong. To be sure, the lowest simple correlation between the intercepts from the unrestricted maximum likelihood procedure and the instrumental variables procedure is .9557, occurring in the third five year period. Similarly, the lowest simple correlation between the unrestricted maximum likelihood procedure and the principal components procedure is .9267 in the first five year period. While even the lowest simple correlations are quite high, the same cannot be said for the rank correlations. The rank correlations between the unrestricted maximum likelihood procedure and the instrumental variables procedure are .9492, .5162, and .5044 in the three five year periods. The corresponding rank correlations with the principal components procedure are .5544, .5379, and .9698. Note that the difference in the magnitudes of the simple and rank correlations is likely to be a reflection of the outlier problem and that the rank correlations on the order of 0.5 suggest considerable differences in the relative rankings of mutual funds by the efficient and the inefficient estimation methods. In short, while the absolute rankings of the funds appear to be relatively insensitive to the estimation method, the relative rankings can be greatly influenced by this choice.

Given the factor loadings and the idiosyncratic variances, the efficiency of the minimum idiosyncratic risk estimator of the basis portfolios increases with the number of securities used in the cross-section. This gain in efficiency, however, is not without cost: namely the increased computational time involved in performing the factor analysis on the greater number of securities. Statistical evidence in Lehmann and Modest(1985b) suggests this gain can be quite substantial. In order to study whether the number of securities used in estimation has an economically significant impact on the performance of the reference portfolios as benchmarks, we performed unrestricted maximum likelihood factor analysis on the first 30, 250, and 750 securities in our randomly sampled data file. Tables 3 and 4 present the evidence on the impact of the number of securities used in estimating the **APT** on inferences regarding mutual fund performance. The tables report results based on a five factor model for security returns which was estimated using the unrestricted maximum likelihood procedure with 30, 250, and 750 securities.[36] The results in Table 3 indicate substantial variation in the mean alphas depending on the number of securities used in the estimation. Previous authors, such as Roll and Ross(1980), have based their inferences concerning the **APT** on maximum likelihood estimation of factor models involving thirty to sixty securities. As is evident, this leads to very different conclusions about the absolute and relative performance of the funds than would be reached from performing the analysis with a much larger number of securities. The difference between the mean alphas using 30 securities and those based on 750 securities is, on an annual basis, +390 basis points, +133 basis points and +233 basis points for the three five year periods respectively. The corresponding differences with the mean intercepts from runs using 250 securities are +138, −90, and +43 basis points. Table 4, as did Table 2 above, presents the simple and rank correlations of the intercepts from the three periods when 30, 250, and 750 securities were used in estimating the factor loadings and idiosyncratic variances of the **APT**. While the correlations are moderately high, they are nowhere near unity. The simple correlations between the intercepts using 30 and 750 securities range from .6513 to .9785, while the corresponding numbers for the rank correlations range from .7314 to .8795. The corresponding range of correlations between the intercepts based on estimation with 250 and 750 securities are .6687 to .9508 for the simple correlations, and .8153 to .9106 for the rank correlations. In short, Tables 1 – 4 suggest that using inefficient estimation procedures due either to an inefficient method or a small number of securities can lead to substantially different conclusions than one would reach with more efficient procedures. Thus the dissimilarities in the relative rankings of the funds found here suggest that the *statistical* differences documented in Lehmann and Modest(1985b) between efficient and inefficient estimation methods can lead to substantively different *economic* conclusions regarding the performance of mutual funds.

In our final comparison of **APT** benchmarks, we examined how sensitive the absolute and relative rankings of the funds are to the presumed number of common factors affecting security returns. In particular, we contrast the performance of basis portfolios constructed under the alternative assumptions that there are five, ten, and fifteen common factors using

[36] Tables $A3$ and $A4$ in the Appendix present the corresponding numbers from a ten factor model of security returns.

the estimated factor loadings and idiosyncratic variances from maximum likelihood factor analysis of 750 securities. Tables 5 and 6 present evidence on whether the number of common factors assumed to be generating security returns has a large impact on the inferences one would make about the absolute and relative rankings of funds. As is readily apparent, the number of common factors assumed to impinge on security returns has far less impact on the rankings of the funds than does the choice of estimation method or number of securities used in estimation. The difference between the mean intercepts from estimating five factors relative to ten factors is (on an annual basis) -2 basis points, $+88$ basis points, and -230 basis points in the three five year periods respectively. The corresponding differences between the mean intercepts using ten and fifteen factors are $+11$, -32 and -1 basis point(s) respectively. Except for the difference between the mean intercepts using five and ten factors in the third and, perhaps, the second five year period, these differences are all quite small. The same picture arises from an examination of the correlations in Table 6. The simple correlations between the intercepts from the five and ten factor models range from .9859 to .9955 and the rank correlations range from .9577 to .9643. The corresponding correlations between the ten and fifteen factor intercepts are also very high with the simple correlations ranging from .9737 to .9904 and the rank correlations ranging from .9785 and .9831. Thus the choice of the number of factors does not appear to be an important one in evaluating the performance of mutual funds. Evidence presented in Lehmann and Modest(1985a) suggests that a ten factor model might be marginally preferable to a five factor model but that there is no apparent advantage to going to fifteen factors. The evidence presented here reinforces the observation that there is no meaningful economic difference in assuming there are ten common sources of systematic risk as opposed to fifteen common factors.

Having compared alternative **APT** benchmarks, the natural question to ask is whether the **APT** has anything different to say about performance evaluation than the **CAPM**. Tables 7 and 8 present summary statistics and correlations that attempt to shed light on this question. As a point of reference we also present summary statistics based on no risk adjustment as well. The difference between the mean intercepts using the **APT** benchmark and the mean intercepts using either of the **CAPM** benchmarks is striking. While the alphas from the **APT** benchmarks are markedly negative in all three periods, the

CAPM alphas are much less negative. A comparison of the average t-statistics suggests that the **CAPM** alphas are not nearly as statistically significant as the **APT** alphas. The means of the **CAPM** alphas using the value-weighted index are (on an annual basis) 353, 567, and 294 basis points higher in the three five year periods than the mean value of the **APT** alphas using the unrestricted maximum likelihood estimation procedure with 750 securities to construct the **APT** benchmark. The corresponding difference between the **CAPM** alphas using the equally weighted index are +483, +584, and −167 basis points. These sharp differences are further indicated by the relatively low simple and rank correlations between the intercepts. The simple correlation between the **APT** intercepts and the **CAPM** intercepts using the value weighted index are .7896, .6545, and .4959. The corresponding rank correlations are .7238, .7076, and .8378. The simple and rank correlations between the intercepts using the equally weighted **CAPM** benchmark and the **APT** benchmark are .6055, .9804, .9003, and .7797, .4860, .7773, respectively. Thus in fact we see that the conclusions one would reach would be dramatically affected by the choice between an **APT** benchmark and a **CAPM** benchmark.

An examination of Table 8 also indicates interesting differences among the two different **CAPM** benchmarks. For instance, the simple correlations between the intercepts constructed using the CRSP value-weighted and equally weighted indices is .9599, .7200, and .4076 in the three periods respectively. The corresponding rank correlations are .5530, .1836, and .6690. It thus appears, unlike the results in Stambaugh(1982), that inferences regarding the relative performance of mutual funds is quite sensitive to the particular market proxy chosen.

What accounts for the sharply negative intercepts? We offer two potential explanations. The first possibility is that there is error in our constructed benchmarks. For example, in Lehmann and Modest(1985a), we found that the **APT** could explain the empirical anomalies involving dividend yield and own variance but could not account for size related anomalies. In particular, we found that the regression of the value weighted **CRSP** index on our basis portfolios yielded significant negative intercepts in each of the five year periods covered by the mutual fund data. The predominantly negative intercepts from the mutual fund regressions could follow from this phenomenon to the extent that funds hold a large part of their portfolios in stocks with large market capitalizations, although it is worth noting that the intercepts from the value weighted regressions were not as large and negative as the mean intercept from the mutual fund regressions. The second explanation involves true or spurious market timing by mutual fund managers. As discussed above in Section III. if the risk of a mutual fund is constant over the sample period. then the fund's alpha is, in principle, an accurate measure of the fund's stock selection ability. However if the fund's risk level is not constant. possibly due to shifts associated with market timing attempts or because of the option nature of levered securities, then the alpha is no longer a measure of the fund's performance ability since the estimated intercept may be arbitrarily positive or negative depending on the covariance between changes in the funds risk posture and the returns on the factors.[37]

We ran quadratic regressions of the mutual funds' returns on the factors and the factors squared, as outlined in equation (27) for the single factor case, to examine whether real or artificial market timing accounts for the incidence of persistently negative intercepts. Under the joint null hypothesis that the risk of the funds is constant and that the returns of the individual securities are generated by a **K** factor linear structure, one should not be able to reject the joint hypothesis that the coefficients on the quadratic terms are zero.[38] If the residuals from the quadratic regression are homoskedastic. then the appropriate joint test is a standard *F-test*. However, if the residuals are heteroskedastic, then the usual *F-statistic* is not appropriate since it fails to possess an *F distribution* in these circumstances and. hence. can lead to incorrect inferences regarding the null hypothesis. Asymptotically, an appropriate test can be conducted using the procedures proposed by Hansen(1982), White(1980) and Hsieh(1983) to construct heteroskedastic-consistent covariance matrices along the lines of the adjusted t-statistics discussed above. This test has the shortcoming that its small sample distribution is not known and reliance must be made on the fact that the test statistic is asymptotically distributed as chi-squared.

In Table 9 we present summary statistics for tests that the risk levels of the funds

[37] See Jagannathan and Korajczyk(1984) for a discussion of the problems associated with the artificial market timing caused by the holding of levered securities. See Pfleiderer and Bhattacharya(1983) and Grinblatt and Titman(1985) for a discussion of the spurious market timing that can arise when managers revise their portfolios more frequently than we observe returns.

[38] This assumes, of course. that we know **K** and that we have basis portfolios which are highly correlated with the factors.

are constant under the assumption that there are five common factors affecting security returns. For each sample period, we provide evidence on the fraction of the funds for which we could reject the hypothesis that the quadratic terms were zero at the 1%, 5%, 10%, and 15% significance levels. We present results from both the F-tests, which are valid under the assumption that the residuals are homoskedastic, and the Chi-squared tests which are asymptotically valid under arbitrary forms of heteroskedasticity. An examination of Table 9 reveals that the Chi-squared tests lead, under the assumption of independence, to a greater number of rejections of the null hypothesis of constant risk levels in all three periods than would have been expected *a priori*. The F-tests on the other hand lead to approximately the number of rejections that would have been expected, except in the second five year period. For instance, in the first five year period, the Chi-square test rejects the null hypothesis that the quadratic terms are zero for 14.6% of the firms at the 1% significance level, 23.8% at the 5% level, 35.4% at the 10% level, and 42.3% at the 15% significance level. The F-test, however, leads to rejection of the null for only 2.3% of the firms at the 1% level, 5.4% at the 5% level, 13.1% at the 10% level, and 20.0% at the 15% significance level. It is difficult to know what to make of these differences as the F-test is only valid under the assumption of homoskedasticity, an assumption which seems to be of dubious validity in this context, and the small sample properties of the Chi-squared test are unknown.[39] We are currently examining the relationship between the intercepts and these test statistics on an individual fund basis to see if this will shed any light on the prevalence of negative alphas.

VI. Conclusion

In Lehmann and Modest(1985b), we provided a comprehensive statistical examination of the merits of different basis portfolio formation strategies. Two of the conclusions which emerged from that study were: (1) that comparatively efficient estimation methods such as maximum likelihood and restricted maximum likelihood factor analysis significantly

[39] One natural way to confront this problem is to carry out a direct test of the homoskedasticity. Monte Carlo evidence presented in Hsieh(1983), however, suggests that direct tests of heteroskedasticity have little power to discriminate between homoskedasticity and heteroskedasticity. Hsieh suggests that even in small samples there is little harm (and potentially a significant gain) to always using the heteroskedastic-consistent standard errors. Unfortunately, this does not speak to the problem here which is the rate of convergence of the test statistic to its asymptotic distribution.

outperform the less efficient instrumental variables and principal components procedures and (2) increasing the the number of securities used to construct the reference portfolios seems to dramatically improve their ability to mimic the common factors. That paper left open. however, the question of whether the use of comparatively inefficient portfolio formation procedures would have a significant quantitative impact on inferences in particular applications such as the evaluation of managed portfolios.

In this paper we have examined the performance of 130 mutual funds over the period January 1968-December 1982 in an attempt find out whether inferences about the performance of these funds are sensitive to the benchmark chosen to measure normal performance. In this regard, we studied the behavior of the intercepts from Jensen-style mutual fund regressions which used different risk measurement procedures. As a consequence, we examined alternative **APT** and **CAPM** benchmarks. In particular, we addressed the question of whether the absolute and relative rankings of the funds depend on the chosen benchmark.

Three conclusions emerged from this comparison. First. absolute and relative mutual fund rankings are quite sensitive to the method used to construct the **APT** benchmark. One would reach very different conclusions about the funds' performance using smaller numbers of securities in the analysis or the less efficient estimation methods than one would arrive at using the maximum likelihood procedures with 750 securities. Second, the rankings of the funds are not very sensitive to the exact number of common sources of systematic risk that are assumed to impinge on security returns. There were virtually no differences in the rankings between the ten and fifteen factor models and only small differences with the five factor benchmark. Third. there is little similarity between the rankings using the standard **CAPM** benchmarks and the **APT** benchmarks which suggests the importance of knowing the appropriate model for risk and expected return in this context. We are currently engaged in research along these lines (Lehmann and Modest(1985a)). Moreover, to the extent that the **CAPM** is the proper theory of expected returns, the results presented here suggest that the choice of the appropriate market proxy is an important one.

In short, the one firm conclusion that can be reached from our analysis is that the choice of what constitutes normal performance is important for evaluating the performance of managed portfolios. It is also worth stressing that these findings are in no way compromised by the potential problems associated with the shifting risk levels of managed funds. These problems only affect the interpretation of the intercepts from the Jensen-style regressions. If the choice of a benchmark were an unimportant one, different benchmarks should have yielded similar results—the overwhelming fact is that they did not.

These findings stand in sharp contrast to much of the conventional wisdom in the literature. We conjecture that many investigators would not have expected substantive differences in the **APT** benchmarks produced by efficient and inefficient estimation methods in our large cross-sections. Conversely, some scholars would doubtless have predicted large differences in the inferences produced by **APT** benchmarks with different numbers of factors. Finally, previous evidence suggests alternative risk adjustment procedures lead to similar inferences in settings other than the present one. Our comprehensive examination of mutual fund performance suggests that each of these intuitions is unreliable in this context.

Along with the three conclusions which have emerged from our analysis, one puzzle has also arisen: the persistent incidence of large and negative alphas. While theoretically it is possible that this negative abnormal performance can be attributed to real or artificial market timing or to a value weighted bias in our constructed benchmarks, the preliminary evidence is not conclusive. We are thus still actively engaged in efforts to explain this phenomenon.

TABLE 1: STATISTICS OF INTERCEPTS ACROSS ESTIMATION METHODS OF THE APT

(standard deviations in parentheses)

Number of factors: 5

Number of funds: 130

Number of Securities Used in Estimation: 750

Sample Period	Statistic	Maximum Likelihood		Restricted Maximum Likelihood		Instrumental Variables		Principal Components	
		Raw Returns	Excess Returns	Raw Returns	Excess Returns	Raw Returns	Excess Returns	Raw Returns	Excess Returns
January 1968-December 1972	Mean	-.00313 (.00431)	-.00406 (.00346)	-.00334 (.00445)	-.00429 (.00354)	-.00286 (.00401)	-.00377 (.00315)	-.00091 (.00356)	-.00234 (.00333)
	Mean Absolute Intercept	0.00394	0.00447	0.00413	0.00468	0.00364	0.00407	0.00258	0.00321
	Average Absolute t-statistic	1.55 (1.20)	1.98 (1.12)	1.61 (1.24)	2.04 (1.14)	1.40 (1.07)	1.83 (1.03)	1.09 (.86)	1.42 (.92)
	Average Absolute t-Adjusted	1.53 (1.22)	2.05 (1.15)	1.59 (1.25)	2.12 (1.17)	1.40 (1.08)	1.91 (1.09)	1.17 (0.93)	1.52 (.98)
January 1973-December 1977	Mean	-.00299 (.00361)	-.00527 (.00338)	-.00295 (.00359)	-.00525 (.00336)	-.00319 (.00348)	-.00540 (.00314)	-.00346 (.00397)	-.00503 (.00359)
	Mean Absolute Intercept	0.00379	0.00555	0.00375	0.00553	0.00381	0.00564	0.00429	0.00539
	Average Absolute t-statistic	1.70 (1.13)	2.60 (1.32)	1.69 (1.12)	2.59 (1.30)	1.85 (1.17)	2.50 (1.14)	2.03 (1.32)	2.41 (1.37)
	Average Absolute t-Adjusted	1.82 (1.21)	2.78 (1.44)	1.80 (1.20)	2.77 (1.42)	2.00 (1.25)	2.61 (1.18)	2.15 (1.39)	2.55 (1.44)
January 1978-December 1982	Mean	-.00074 (.00269)	-.00126 (.00248)	-.00071 (.00267)	-.00122 (.00248)	-.00042 (.00270)	-.00074 (.00270)	-.00042 (.00271)	-.00117 (.00243)
	Mean Absolute Intercept	0.00210	0.00212	0.00208	0.00209	0.00203	0.00206	0.00202	0.00206
	Average Absolute t-statistic	0.89 (.75)	0.94 (.79)	0.88 (.74)	0.92 (.78)	0.80 (.65)	0.83 (.70)	0.82 (.69)	0.87 (.71)
	Average Absolute t-Adjusted	0.96 (.81)	1.02 (.86)	0.95 (.80)	1.00 (.85)	0.85 (.69)	0.88 (.74)	0.89 (.79)	0.92 (.76)

TABLE 2: CORRELATIONS OF INTERCEPTS ACROSS ESTIMATION METHODS OF THE APT
(Excess Return Version of APT)

Number of factors: 5
Number of funds: 130
Number of Securities Used in Estimation: 750

Sample Period		SIMPLE CORRELATIONS				RANK CORRELATIONS			
		Maximum Likelihood	Restricted Maximum Likelihood	Instrumental Variable	Principal Components	Maximum Likelihood	Restricted Maximum Likelihood	Instrumental Variables	Principal Components
January 1968-December 1972	Maximum Likelihood	1.0000	.9993	.9640	.9267	1.0000	.9999	.9492	.5544
	Restricted Maximum Likelihood	.9993	1.0000	.9931	.9144	.9999	1.0000	.9491	.5516
	Instrumental Variables	.9640	.9590	1.0000	.9122	.9492	.9491	1.0000	.5085
January 1973-December 1977	Maximum Likelihood	1.0000	.9999	.9908	.9926	1.0000	.9999	.5162	.5379
	Restricted Maximum Likelihood	.9999	1.0000	.9914	.9919	.9999	1.0000	.5165	.5373
	Instrumental Variables	.9908	.9919	1.0000	.9752	.5162	.5165	1.0000	.3950
January 1978-December 1982	Maximum Likelihood	1.0000	.9996	.9557	.9800	1.0000	.9999	.5044	.9698
	Restricted Maximum Likelihood	.9996	1.0000	.9576	.9784	.9999	1.0000	.5041	.9702
	Instrumental Variables	.9557	.9576	1.0000	.9226	.5044	.5041	1.0000	.4987

TABLE 3: STATISTICS OF INTERCEPTS ACROSS NUMBER OF SECURITIES USED IN ESTIMATING THE APT
(standard deviations in parentheses)

Number of factors: 5
Estimation method: Maximum Likelihood
Number of funds: 130

Sample Period	Statistic	30		250		750	
		Raw Returns	Excess Returns	Raw Returns	Excess Returns	Raw Returns	Excess Returns
January 1968-December 1972	Mean	0.00052 (.00382)	-.00087 (.00332)	-.00157 (.00363)	-.00292 (.00332)	-.00313 (.00431)	-.00406 (.00346)
	Mean Absolute Intercept	0.00276	0.00241	0.00279	0.00344	0.00394	0.00447
	Average Absolute t-statistic	0.96 (.79)	0.76 (.62)	0.99 (.82)	1.39 (.91)	1.55 (1.20)	1.98 (1.12)
	Average Absolute t-Adjusted	1.01 (.84)	0.80 (.65)	1.04 (.85)	1.46 (.95)	1.53 (1.22)	2.05 (1.15)
January 1973-December 1977	Mean	-.00194 (.00379)	-.00417 (.00358)	-.00301 (.00354)	-.00602 (.00333)	-.00299 (.00361)	-.00527 (.00338)
	Mean Absolute Intercept	0.00335	0.00462	0.00376	0.00622	0.00379	0.00555
	Average Absolute t-statistic	1.09 (.76)	1.13 (.75)	1.60 (1.00)	2.53 (1.10)	1.70 (1.13)	2.60 (1.32)
	Average Absolute t-Adjusted	1.21 (.86)	1.27 (.84)	1.77 (1.13)	2.57 (1.11)	1.82 (1.21)	2.78 (1.44)
January 1978-December 1982	Mean	0.00227 (.00293)	0.00065 (.00302)	-.00031 (.00267)	-.00091 (.00250)	-.00074 (.00269)	-.00127 (.00248)
	Mean Absolute Intercept	0.00293	0.00219	0.00195	0.00196	0.00210	0.00212
	Average Absolute t-statistic	0.96 (.70)	0.63 (.57)	0.74 (.70)	0.75 (.66)	0.89 (.75)	0.94 (.79)
	Average Absolute t-Adjusted	1.00 (.71)	0.70 (.62)	0.81 (.71)	0.81 (.71)	0.96 (.81)	1.02 (.86)

TABLE 4: CORRELATIONS OF INTERCEPTS ACROSS NUMBER OF SECURITIES USED IN ESTIMATING THE APT
(Excess Return Version of APT)

Number of factors: 5
Estimation Method: Maximum Likelihood
Number of funds: 130

		SIMPLE CORRELATIONS			RANK CORRELATIONS		
Sample Period		30	250	750	30	250	750
January 1968-December 1972	30	1.0000	.8498	.7577	1.0000	.8936	.8669
	250	.8498	1.0000	.9706	.8936	1.0000	.9704
January 1973-December 1977	30	1.0000	.9508	.9785	1.0000	.8153	.7314
	250	.9508	1.0000	.9911	.8153	1.0000	.9519
January 1978-December 1982	30	1.0000	.6687	.6513	1.0000	.9106	.8795
	250	.6687	1.0000	.9755	.9106	1.0000	.9572

TABLE 5: STATISTICS OF INTERCEPTS ACROSS NUMBER OF FACTORS
(standard deviations in parentheses)

Estimation method: Maximum Likelihood
Number of Securities Used in Estimation: 750
Number of funds: 130

Sample Period	Statistic	5		10		15	
		Raw Returns	Excess Returns	Raw Returns	Excess Returns	Raw Returns	Excess Returns
January 1968-December 1972	Mean	-.00313 (.00431)	-.00406 (.00346)	-.00268 (.00399)	-.00404 (.00321)	-.00357 (.00488)	-.00395 (.00377)
	Mean Absolute Intercept	0.00394	0.00447	0.00351	0.00439	0.00425	0.00438
	Average Absolute t-statistic	1.55 (1.20)	1.98 (1.12)	1.34 (1.04)	2.05 (1.16)	1.45 (1.14)	1.88 (1.17)
	Average Absolute t-Adjusted	1.53 (1.21)	2.05 (1.15)	1.51 (1.13)	2.30 (1.33)	1.74 (1.33)	2.22 (1.38)
January 1973-December 1977	Mean	-.00299 (.00361)	-.00527 (.00338)	-.00223 (.00313)	-.00454 (.00300)	-.00250 (.00316)	-.00481 (.00318)
	Mean Absolute Intercept	0.00379	0.00555	0.00298	0.00483	0.00316	0.00511
	Average Absolute t-statistic	1.70 (1.13)	2.60 (1.32)	1.40 (1.01)	2.37 (1.17)	1.46 (1.02)	2.52 (1.21)
	Average Absolute t-Adjusted	1.82 (1.21)	2.78 (1.44)	1.59 (1.17)	2.68 (1.35)	1.68 (1.19)	2.89 (1.41)
January 1978-December 1982	Mean	-.00074 (.00269)	-.00127 (.00248)	-.00271 (.00279)	-.00321 (.00275)	-.00293 (.00292)	-.00328 (.00281)
	Mean Absolute Intercept	0.00210	0.00212	0.00306	0.00342	0.00329	0.00352
	Average Absolute t-statistic	0.89 (.75)	0.94 (.79)	1.26 (.87)	1.45 (.92)	1.38 (.93)	1.48 (.94)
	Average Absolute t-Adjusted	0.96 (.81)	1.02 (.86)	1.51 (1.05)	1.72 (1.09)	1.73 (1.22)	1.86 (1.19)

TABLE 6: CORRELATIONS OF INTERCEPTS ACROSS NUMBER OF FACTORS
(Excess Return Version of APT)

Estimation Method: Maximum Likelihood
Number of Securities Used in Estimation: 750
Number of funds: 130

Sample Period		Simple Correlations			Rank Correlations		
		5	10	15	5	10	15
January 1968-December 1972	5	1.0000	.9955	.9750	1.0000	.9577	.9370
	10	.9955	1.0000	.9737	.9577	1.0000	.9831
January 1973-December 1977	5	1.0000	.9859	.9732	1.0000	.9643	.9469
	10	.9859	1.0000	.9904	.9643	1.0000	.9785
January 1978-December 1982	5	1.0000	.9955	.9750	1.0000	.9643	.9469
	10	.9955	1.0000	.9737	.9643	1.0000	.9785

TABLE 7: STATISTICS OF INTERCEPTS ACROSS BENCHMARKS

APT: Number of factors: 5
Estimation method: Maximum Likelihood
Number of Securities Used in Estimation: 750
Number of funds: 130
(standard deviations in parentheses)

Sample Period	Statistic	APT Raw Returns	APT Excess Returns	CAPM Value Weighted Excess Returns	CAPM Equally Weighted Excess Returns	No Risk Adjustment Excess Returns
January 1968- December 1972	Mean	-.00313 (.00431)	-.00406 (.00346)	-.00117 (.00365)	-.00012 (.00352)	0.00545 (.00339)
	Mean Absolute Intercept	0.00394	0.00447	0.00260	0.00234	0.00576
	Average Absolute t-statistic	1.55 (1.20)	1.98 (1.12)	0.99 (.81)	0.81 (.71)	1.03 (.58)
	Average Absolute t-Adjusted	1.53 (1.22)	2.05 (1.15)	1.01 (.82)	0.83 (.72)	1.04 (.58)
January 1973- December 1977	Mean	-.00299 (.00361)	-.00527 (.00338)	-.00066 (.00378)	-.00053 (.00410)	0.00111 (.00402)
	Mean Absolute Intercept	.00379	0.00555	0.00302	0.00569	0.00318
	Average Absolute t-statistic	1.70 (1.13)	2.60 (1.32)	1.31 (.87)	1.52 (.82)	0.53 (.53)
	Average Absolute t-Adjusted	1.82 (1.21)	2.78 (1.44)	1.31 (.86)	1.58 (.86)	0.54 (.53)
January 1978- December 1982	Mean	-.00074 (.00269)	-.00126 (.00248)	0.00116 (.00332)	-.00266 (.00273)	0.01363 (.00401)
	Mean Absolute Intercept	0.00210	0.00212	0.00245	0.00329	0.01363
	Average Absolute t-statistic	0.89 (.75)	0.94 (.79)	0.91 (.69)	1.14 (.69)	2.11 (.45)
	Average Absolute t-Adjusted	0.96 (.81)	1.02 (.86)	0.94 (.71)	1.18 (.71)	2.12 (.45)

TABLE 8: STATISTICS OF INTERCEPTS ACROSS BENCHMARKS

APT: Number of factors: 5
Estimation method: Maximum Likelihood
Number of Securities Used in Estimation: 750
Number of funds: 130
Excess Return Version of APT

		SIMPLE CORRELATIONS				RANK CORRELATIONS			
			CAPM				CAPM		
Sample Period		APT	Value Weighted	Equally Weighted	No Risk Adjustment	APT	Value Weighted	Equally Weighted	No Risk Adjustment
	APT	1.0000	.7896	.6055	-.3436	1.0000	.7238	.7797	-.0819
January 1968-December 1972	CAPM Value Weighted	.7896	1.0000	.9599	.2355	.7238	1.0000	.5530	.0376
	CAPM Equally Weighted	.6055	.9599	1.0000	.4952	.7797	.5530	1.0000	-.1132
	APT	1.0000	.6545	.9804	.2723	1.0000	.7076	.4860	-.0181
January 1973-December 1977	CAPM Value Weighted	.6545	1.0000	.7200	.8876	.7076	1.0000	.1836	-.2084
	CAPM Equally Weighted	.9804	.7200	1.0000	.3776	.4860	.1836	1.0000	.1762
	APT	1.0000	.4959	.9003	-.2715	1.0000	.8378	.7773	-.0898
January 1978-December 1982	CAPM Value Weighted	.4959	1.0000	.4076	.5755	.8378	1.0000	.6690	.0861
	CAPM Equally Weighted	.9003	.4076	1.0000	-.4952	.7773	.6690	1.0000	.2721

TABLE A1: STATISTICS OF INTERCEPTS ACROSS ESTIMATION METHODS OF THE APT
(standard deviations in parentheses)

Number of factors: 10
Number of funds: 130
Number of Securities Used in Estimation: 750

Sample Period	Statistic	Maximum Likelihood		Restricted Maximum Likelihood		Instrumental Variables		Principal Components	
		Raw Returns	Excess Returns	Raw Returns	Excess Returns	Raw Returns	Excess Returns	Raw Returns	Excess Returns
January 1968-December 1972	Mean	-.00268 (.00399)	-.00404 (.00321)	-.00294 (.00415)	-.00435 (.00327)	-.00398 (.00452)	-.00427 (.00339)	-.00122 (.00336)	-.00289 (.00301)
	Average Absolute Intercept	0.00351	0.00439	0.00373	0.00468	0.00462	0.00467	0.00264	0.00347
	Average Absolute t-statistic	1.34 (1.04)	2.05 (1.16)	1.38 (1.05)	2.15 (1.19)	1.67 (1.13)	1.98 (1.05)	1.11 (.85)	1.53 (.96)
	Average Absolute t-Adjusted	1.51 (1.13)	2.30 (1.33)	1.56 (1.15)	2.43 (1.36)	1.79 (1.17)	2.20 (1.22)	1.25 (.95)	1.71 (1.09)
January 1973-December 1977	Mean	-.00223 (.00313)	-.00454 (.00300)	-.00215 (.00311)	-.00446 (.00297)	-.00337 (.00351)	-.00582 (.00324)	-.00331 (.00369)	-.00544 (.00344)
	Average Absolute Intercept	0.00298	0.00483	0.00291	0.00475	0.00396	0.00601	0.00407	0.00570
	Average Absolute t-statistic	1.40 (1.01)	2.36 (1.17)	1.36 (.99)	2.29 (1.14)	1.71 (1.07)	2.72 (1.19)	1.76 (1.10)	2.37 (1.19)
	Average Absolute t-Adjusted	1.59 (1.17)	2.68 (1.35)	1.54 (1.15)	2.60 (1.31)	1.95 (1.24)	3.04 (1.38)	1.99 (1.24)	2.63 (1.33)
January 1978-December 1982	Mean	-.00271 (.00279)	-.00321 (.00275)	-.00280 (.00288)	-.00345 (.00294)	-.00124 (.00264)	-.00163 (.00261)	-.00047 (.00266)	-.00148 (.00236)
	Average Absolute Intercept	0.00306	0.00343	0.00314	0.00366	0.00223	0.00234	0.00201	0.00220
	Average Absolute t-statistic	1.26 (.87)	1.45 (.92)	1.26 (.88)	1.50 (.94)	0.92 (.73)	0.99 (.77)	0.83 (.66)	0.91 (.69)
	Average Absolute t-Adjusted	1.50 (1.05)	1.72 (1.09)	1.51 (1.07)	1.79 (1.12)	1.07 (.86)	1.16 (.95)	0.94 (.75)	1.04 (.78)

TABLE A2: CORRELATIONS OF INTERCEPTS ACROSS ESTIMATION METHODS OF THE APT
(Excess Return Version of APT)

Number of factors: 10
Number of funds: 130
Number of Securities Used in Estimation: 750

Sample Period		SIMPLE CORRELATIONS				RANK CORRELATIONS			
		Maximum Likelihood	Restricted Maximum Likelihood	Instrumental Variable	Principal Components	Maximum Likelihood	Restricted Maximum Likelihood	Instrumental Variables	Principal Components
January 1968- December 1972	Maximum Likelihood	1.0000	.9989	.9814	.9547	1.0000	.9998	.9374	.6141
	Restricted Maximum Likelihood	.9989	1.0000	.9801	.9465	.9998	1.0000	.9366	.6124
	Instrumental Variables	.9814	.9801	1.0000	.9423	.9374	.9366	1.0000	.5228
January 1973- December 1977	Maximum Likelihood	1.0000	.9996	.9872	.9787	1.0000	.9999	.6149	.5170
	Restricted Maximum Likelihood	.9996	1.0000	.9870	.9744	.9999	1.0000	.6142	.5164
	Instrumental Variables	.9872	.9870	1.0000	.9710	.6149	.6142	1.0000	.4330
January 1978- December 1982	Maximum Likelihood	1.0000	.9969	.8604	.8044	1.0000	.9994	.5062	.9468
	Restricted Maximum Likelihood	.9969	1.0000	.8280	.7656	.9994	1.0000	.5035	.9459
	Instrumental Variables	.8604	.8280	1.0000	.8863	.5062	.5035	1.0000	.4982

TABLE A3: STATISTICS OF INTERCEPTS ACROSS NUMBER OF SECURITIES USED IN ESTIMATING THE APT
(standard deviations in parentheses)

Number of factors: 10
Estimation method: Maximum Likelihood
Number of funds: 130

Sample Period	Statistic	30 Raw Returns	30 Excess Returns	250 Raw Returns	250 Excess Returns	750 Raw Returns	750 Excess Returns
January 1968-December 1972	Mean	0.00185 (.00387)	-.00087 (.00332)	-.00152 (.00396)	-.00292 (.00339)	-.00268 (.00399)	-.00406 (.00321)
	Average Absolute Intercept	0.00318	0.00241	0.00284	0.00344	0.00351	0.00447
	Average Absolute t-statistic	1.09 (.83)	0.76 (.59)	1.02 (.84)	1.39 (1.01)	1.34 (1.04)	1.98 (1.16)
	Average Absolute t-Adjusted	1.22 (.96)	0.80 (.64)	1.15 (.90)	1.46 (1.08)	1.51 (1.13)	2.05 (1.33)
January 1973-December 1977	Mean	-.00340 (.00448)	-.00614 (.00375)	-.00317 (.00354)	-.00652 (.00346)	-.00223 (.00313)	-.00454 (.00300)
	Average Absolute Intercept	0.00463	0.00639	0.00386	0.00667	0.00298	0.00483
	Average Absolute t-statistic	1.62 (.99)	1.87 (1.07)	1.63 (1.02)	2.76 (1.17)	1.40 (1.01)	2.36 (1.17)
	Average Absolute t-Adjusted	1.82 (1.15)	2.04 (1.19)	1.89 (1.21)	2.91 (1.24)	1.59 (1.17)	2.68 (1.35)
January 1978-December 1982	Mean	0.00182 (.00332)	0.00087 (.00321)	-.00167 (.00273)	-.00269 (.00260)	-.00271 (.00279)	-.00321 (.00275)
	Average Absolute Intercept	0.00299	0.00246	0.00245	0.00302	0.00306	0.00343
	Average Absolute t-statistic	0.95 (.70)	0.70 (.55)	0.89 (.69)	1.12 (.75)	1.26 (.87)	1.45 (.92)
	Average Absolute t-Adjusted	1.07 (.78)	0.81 (.65)	1.02 (.84)	1.31 (.89)	1.51 (1.05)	1.72 (1.09)

TABLE A4: CORRELATIONS OF INTERCEPTS ACROSS NUMBER OF SECURITIES USED IN ESTIMATING THE APT
(Excess Return Version of APT)

Number of factors: 10
Estimation Method: Maximum Likelihood
Number of funds: 130

		SIMPLE CORRELATIONS			RANK CORRELATIONS		
Sample Period		30	250	750	30	250	750
January 1968-December 1972	30	1.0000	.5920	.6137	1.0000	.8618	.8348
	250	.5920	1.0000	.9769	.8618	1.0000	.9605
January 1973-December 1977	30	1.0000	.9813	.9708	1.0000	.6908	.7463
	250	.9813	1.0000	.9836	.6908	1.0000	.9047
January 1978-December 1982	30	1.0000	.2403	.0903	1.0000	.8742	.8468
	250	.2403	1.0000	.9612	.8742	1.0000	.9549

References

ADMATI, ANAT R. AND PAUL C. PFLEIDERER [1985], "Interpreting the Factor Risk Premia in the Arbitrage Pricing Theory," *Journal of Economic Theory*, Vol. 35, pp. 191-195.

ADMATI, ANAT R. AND STEPHEN A. ROSS [1985], "Measuring Investment Performance in a Rational Expectations Model," *Journal of Business*, Vol. 58, pp. 1-26.

ARNOTT, ROBERT D. [1980], "Cluster Analysis and Stock Price Movements," *Financial Analysts Journal*, Vol. 36, pp. 56-62.

BANZ, ROLF [1981], "The Relationship Between Return and Market Value of Common Stocks," *Journal of Financial Economics*, Vol. 9, pp. 3-18.

BASU, S. [1977], "Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of Market Efficiency," *Journal of Finance*, Vol. 32, pp. 663-682.

BLACK, FISCHER, MICHAEL C. JENSEN AND MYRON SCHOLES [1972], "The Capital Asset Pricing Model: Some Empirical Tests." In Michael C. Jensen, *Studies in the Theory of Capital Markets*, New York: Praeger.

BLACK, FISCHER AND MYRON SCHOLES [1974], "The Effects of Dividend Yields and Dividend Policy on Common Stock Prices and Returns," *Journal of Financial Economics*, Vol. 1, pp. 1-22.

BLUME, MARSHALL E. AND IRWIN FRIEND [1973], "A New Look at the Capital Asset Pricing Model," *Journal of Finance*, Vol. 28, pp. 19-34.

BLUME, MARSHALL E. AND ROBERT F. STAMBAUGH [1983], "Biases in Computed Returns: An Application to the Size Effect," *Journal of Financial Economics*, Vol. 12, pp. 387-404.

BREEN, BILL, RAVI JAGANNATHAN AND RONI OFFER [1984], "Market Timing and Investment Performance - A Comment," Working Paper 111, Banking Research Center, Kellogg Graduate School of Management, Northwestern University.

BROWN, STEPHEN J. AND MARK T. WEINSTEIN [1983], "A New Approach to Testing Arbitrage Pricing Models: The Bilinear Paradigm," *Journal of Finance*, Vol. 28, pp. 711-743.

CANNISTRARO, RICHARD S. [1973], "The Predictive Ability of Price/Earnings Ratios for Determining Risk-Adjusted Performance of Common Stocks," Unpublished, Sloan School of Management, MIT.

CHAMBERLAIN, GARY [1983], "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica*, Vol. 51, pp. 1305-1323.

CHAMBERLAIN, GARY AND MICHAEL ROTHSCHILD [1983], "Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, Vol. 51, pp. 1281-1304.

CHEN, NAI-FU [1983], "Some Empirical Tests of the Theory of Arbitrage Pricing," *Journal of Finance*, Vol. 38, pp. 1393-1414.

CHEN, NAI-FU AND JONATHAN E. INGERSOLL JR. [1983], "Exact Pricing in Linear Factor Models with Infinitely Many Assets: A Note," *Journal of Finance*, Vol. 38, pp. 985-988.

CONNOR, GREGORY [1984], "A Unified Beta Pricing Theory," *Journal of Economic Theory*, Vol. 34, pp. 13-31.

————[1985], "An Intertemporal Equilibrium Beta Pricing Model," Unpublished, Kellogg Graduate School of Management, Northwestern University.

CONNOR, GREGORY AND ROBERT A. KORAJCZYK [1984], "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," Unpublished, Kellogg Graduate School of Management, Northwestern University.

DEMPSTER, ARTHUR P., N.M. LAIRD AND DONALD B. RUBIN [1977], "Maximum Likelihood from Incomplete Data via the E-M Algorithm," *Journal of the Royal Statistical Society*, Vol. Series B 39, pp. 1-38.

DHRYMES, PHOEBUS J., IRWIN FRIEND, MUSTAFA M. GULTEKIN AND N. BULENT GULTEKIN [1985], "An Empirical Examination of the Implications of Arbitrage Pricing Theory," *Journal of Banking and Finance*, Vol. 9, pp. 73-99.

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*, Vol. 39, pp. 323-346.

DYBVIG, PHILIP H. [1983], "An Explicit Bound on Deviations from APT Pricing in a Finite Economy," *Journal of Financial Economics*, Vol. 12, pp. 483-496.

DYBVIG, PHILIP H. AND STEPHEN A. ROSS [1981], "The Simple Analytics of Performance Measurement," Unpublished, Yale University.

————[1983], "Yes, The APT Is Testable," Unpublished, Yale University.

FAMA, EUGENE F. [1984], "Term Premiums in Bond Returns," *Journal of Financial Economics*, Vol. 13, pp. 529-546.

FAMA, EUGENE F. AND JAMES D. MACBETH [1973], "Risk, Return, and Equilibrium: Some Empirical Tests," *Journal of Political Economy*, Vol. 81, pp. 607-636.

FARRELL, J.L. [1974], "Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings," *Journal of Business*, Vol. 47, pp. 186-207.

FEENEY, GEORGE J. AND DONALD D. HESTER [1967], "Stock market Indices: A Principal Components Analysis." In Donald D. Hester and James Tobin, *Risk Aversion and Portfolio Choice*, New York: John Wiley and Sons.

FIRTH, MICHAEL A. [1977], "The Investment Performance of Unit Trusts in the Period 1965-1975," *Journal of Money, Credit and Banking*, Vol. 7, pp. 311-333.

FRIEND, IRWIN, MARSHALL E. BLUME AND JEAN CROCKETT [1970], *Mutual Funds and Other Institutional Investors: A New Perspective*, New York: McGraw-Hill.

GEHR JR., ADAM [1975], "Some Tests of the Arbitrage Pricing Theory," *Journal of the Midwest Finance Association*, Vol. , pp. 91-105.

GIBBONS, MICHAEL R. [1982], "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, Vol. 10, pp. 3-27.

——————[1983], "Empirical Examination of the Return Generating Process of the Arbitrage Pricing Theory," Unpublished, Graduate School of Business, Stanford University.

GRINBLATT, MARK AND SHERIDAN TITMAN [1983], "Factor Pricing in a Finite Economy," *Journal of Financial Economics*, Vol. 12, pp. 495-507.

——————[1984], "The Relationship Between Mean-Variance Efficiency and Arbitrage Pricing," Working Paper 4-83, Graduate School of Management, UCLA.

——————[1985], "Portfolio Performance Evaluation: Old Issues and New Insights," Working Paper 16-84, Graduate School of Management, UCLA.

HAGGLUND, GOSTA [1982], "Factor Analysis by Instrumental Variables Methods," *Psychometrika*, Vol. 47, pp. 209-221.

HANSEN, LARS P. [1982], "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, pp. 1029-1054.

HSIEH, DAVID A. [1983], "A Heteroscedasticity-Consistent Covariance Matrix Estimator for Time Series Regressions," *Journal of Econometrics*, Vol. 22, pp. 281-290.

HUBERMAN, GUR [1982], "A Simple Approach to Arbitrage Pricing Theory," *Journal of Economic Theory*, Vol. 28, pp. 183-191.

HUGHES, PATRICIA [1982], "A Test of the Arbitrage Pricing Theory," Unpublished, University of British Columbia.

INGERSOLL JR., JONATHAN E. [1984], "Some Results in the Theory of Arbitrage Pricing," *Journal of Finance*, Vol. 39, pp. 1021-1039.

JAGANNATHAN, RAVI AND ROBERT A. KORAJCZYK [1984], "Assessing The Market Timing Performance of Managed Portfolios," Working Paper 118, Banking Research Center, Kellogg Graduate School of Management, Northwestern University.

JENSEN, MICHAEL C. [1968], "The Performance of Mutual Funds in the Period 1945-1964," *Journal of Finance*, Vol. 23, pp. 389-416.

——————[1969], "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," *Journal of Business*, Vol. 42, pp. 167-247.

——————[1972], "Optimal Utilization of Market Forecasts and the Evaluation of Investment Portfolio Performance." In G. P. Szego and Karl Shell, *Mathematical Methods in Investment and Finance*, Amsterdam: North Holland.

JOBSON, J.D. AND BOB M. KORKIE [1982], "Potential Performance and Tests of Portfolio Efficiency," *Journal of Finincial Economics*, Vol. 10, pp. 433-466.

KING, BENJAMIN F. [1966], "Market and Industry Factors in Stock Price Behavior," *Journal of Business*, Vol. 39, pp. 139-190.

LAWLEY, D.N. AND A.E. MAXWELL [1971], *Factor Analysis as a Statistical Method*, New York: Elsevier.

LEHMANN, BRUCE N. AND DAVID M. MODEST [1985a], "The Empirical Foundations of the Arbitrage Pricing Theory I: The Empirical Tests," Unpublished, Graduate School of Business, Columbia University.

——————[1985b], "The Empirical Foundations of the Arbitrage Pricing Theory II: The Optimal Construction of Basis Portfolios," Unpublished, Graduate School of Business, Columbia University.

——————[1985c], "Temporal Aggregation and the Optimal Construction of Basis Portfolios," Unpublished, Graduate School of Business, Columbia University.

LINTNER, JOHN [1965], "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, Vol. 47, pp. 13-37.

LITZENBERGER, ROBERT H. AND KRISHNA RAMASWAMY [1979], "The Effects of Dividends on Common Stock Prices: Theory and Empirical Evidence," *Journal of Financial Economics*, Vol. 7, pp. 163-195.

MADANSKY, ALBERT [1964], "Instrumental Variables in Factor Analysis," *Psychometrika*, Vol. 29, pp. 105-113.

MADANSKY, ALBERT AND TERRY MARSH [1985], "The Factor Structure of Asset Returns: How Many Factors Are There and Which Ones Are Priced," Unpublished, Sloan School of Management, M.I.T.

MAINS, NORMAN E. [1977], "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios: Comment," *Journal of Business*, Vol. 50, pp. 371-384.

MARKOWITZ, HARRY [1952], "Portfolio Selection," *Journal of Finance*, Vol. 7, pp. 77-91.

MAYERS, DAVID AND EDWARD RICE [1979], "Measuring Portfolio Performance and Empirical Content of Asset Pricing Models," *Journal of Financial Economics*, Vol. 7, pp. 3-29.

MCDONALD, JOHN G. [1973], "French Mutual fund Performance: Evaluation of Internationally Diversified Portfolios," *Journal of Finance*, Vol. 28, pp. 1161-1180.

——————[1974], "Objectives and Performance of Mutual Funds, 1960-1969," *Journal of Financial and Quantitative Analysis*, Vol. 9, pp. 311-333.

MILLER, RUPERT G. [1966], *Simultaneous Statistical Inference*, New York: McGraw-Hill.

MOSSIN, JAN [1966], "Equilibrium in a Capital Asset Market," *Econometrica*, Vol. 34, pp. 768-783.

PFLEIDERER, PAUL C. AND SUDIPTO BHATTARCHARYA [1983], "A Note on Performance Evaluation," Unpublished, Stanford University and the University of California at Berkeley.

PFLEIDERER, PAUL C. AND PETER REISS [1983], "Testing the APT with Arbitrage Portfolios," Unpublished, Graduate School of Business, Stanford University.

REINGANUM, MARC R. [1981], "Misspecification of Capital Asset Pricing: Empirical Anomolies Based on Earnings Yields and Market Values," *Journal of Financial Economics*, Vol. 9, pp. 19-46.

———[1981], "The Arbitrage Pricing Theory: Some Empirical Results," *Journal of Finance*, Vol. 36, pp. 313-321.

ROLL, RICHARD W. [1977], "A Critique of the Asset Pricing Theory's Tests-Part I: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, Vol. 4, pp. 129-176.

———[1978], "Ambiguity When Performance Is Measured By The Securities Market Line," *Journal of Finance*, Vol. 33, pp. 1051-69.

———[1983], "On Computing Mean Returns and the Small Firm Premium," *Journal of Financial Economics*, Vol. 12, pp. 371-386.

ROLL, RICHARD W. AND STEPHEN A. ROSS [1980], "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, Vol. 35, pp. 1073-1103.

———[1984], "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory: A Reply," *Journal of Finance*, Vol. 39, pp. 347-350.

ROSENBERG, BARR AND VINAY MARATHE [1979], "Tests of Capital Asset Pricing Hypotheses." In , *Research in Finance: A Research Annual*, Greenwich, Connecticut: JAI Press.

ROSS, STEPHEN A. [1976], "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, Vol. 13, pp. 341-60.

———[1977], "Risk,Return, and Arbitrage." In Irwin Friend and James L. Bicksler, *Risk Return in Finance*, Cambridge, Massachusetts: Ballinger.

SHANKEN, JAY [1982], "The Arbitrage Pricing Theory: Is It Testable?," *Journal of Finance*, Vol. 37, pp. 1129-40.

———[1984], "Multivariate Tests of the Zero-Beta CAPM." Unpublished, Graduate School of Business Administration, University of California Berkeley.

SHARPE, WILLIAM F. [1963], "A Simplified Model for Portfolio Analysis," *Management Science*, Vol. 9, pp. 277-93.

———[1964], "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, Vol. 19, pp. 425-42.

———[1966], "Mutual Fund Performance," *Journal of Business*, Vol. 39, pp. 119-138.

———[1967], "Linear Programming Algorithm for Mutual Fund Portfolio Selection," *Management Science*, Vol. 13, pp. 499-510.

STAMBAUGH, ROBERT F. [1982], "On the Exclusion of Assets from Tests of the Two Parameter Model," *Journal of Financial Economics*, Vol. 10, pp. 235-68.

TREYNOR, JACK L. [1965], "How to Rate Management of Investment Funds," *Harvard Business Review*, Vol. 43, pp. 63-75.

TREYNOR, JACK L. AND F. MAZUY [1966], "Can Mutual Funds Outguess the Market," *Harvard Business Review*, Vol. 44, pp. 131-136.

VERRECHIA, ROBERT [1980], "The Mayers-Rice Conjecture: A Counterexapmle," *Journal of Financial Economics*, Vol. 8, pp. 87-100.

WHITE, HALBERT [1980], "A Heteroscedasticity-Consistent Covariance Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, Vol. 48, pp. 817-38.