NBER WORKING PAPER SERIES

THE EMPIRICAL FOUNDATIONS OF THE ARBITRAGE PRICING THEORY I: THE EMPIRICAL TESTS

Bruce N. Lehmann

David M. Modest

Working Paper No. 1725

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

We are grateful to the Faculty Research Fund of the Columbia Business School and the Institute for Quantitative Research in Finance for their support. 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 #1725
October 1985

The Empirical Foundations of the Arbitrage Pricing Theory I: The Empirical Tests

ABSTRACT

This paper provides a detailed and extensive examination of the validity of the APT based on maximum likelihood factor analysis of large cross-sections of securities. Our empirical implementation of the theory proved incapable of explaining expected returns on portfolios composed of securities with different market capitalizations although it provided an adequate account of the expected returns of portfolios formed on the basis of dividend yield and own variance where risk adjustment with the CAPM employing the usual market proxies failed. In addition, it appears that the zero beta version of the APT is sharply rejected in favor of the riskless rate model and that there is little basis for discriminating among five and ten factor versions of the theory.

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

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

I. Introduction

The last thirty years have witnessed great changes in the field of finance associated with the emergence of modern portfolio theory. A profession once populated with institutionally oriented students has been transformed into one dominated by scholars with a more scientific orientation. The Capital Asset Pricing Model (**CAPM**) was an outgrowth of this new emphasis and has served as the basis of modern portfolio theory for much of this period. The early tests of the **CAPM** (implemented with the equally-weighted index as a proxy for the market portfolio) revealed few violations of the zero beta model which could not be ascribed to statistical problems. In particular, the only reliably documented evidence concerned the slope of the empirical security market line which appeared shallower than that predicted by the theory. The tentative success of these initial investigations provided no hint of the disastrous empirical failure which followed: a large body of persuasive evidence which suggested that the dividend yields, market capitalizations, and price-earnings ratios of common stocks were strongly related to expected returns after risk adjustments based on the equally-weighted or value-weighted indices.

Not surprisingly, financial economists have responded to this surfeit of evidence regarding the inefficiency of the usual market proxies by considering aspects of the economic environment omitted in the static **CAPM**. Chief among these are models which incorporate intertemporal fluctuations in investment opportunities, *a priori* restrictions on the covariance structure of security returns, and institutional characteristics such as taxation. Empirical investigations of these alternative approaches have resulted in few firm conclusions regarding their comparative merits. In particular, none of these theories has yet proved fully consistent with the collection of empirical regularities alluded to above.[1] Moreover, Roll(1977) questioned the scientific relevance of such findings for the **CAPM** which implies

[1] The main reference on the impact of dividend taxation on asset pricing is Litzenberger and Ramaswamy(1979) which has been criticized in Miller and Scholes(1978,1982). It seems clear from the evidence presented in Blume(1979) and in Elton, Gruber, and Rentzler(1983) regarding the behavior of zero dividend stocks that the dividend effect is not exclusively a tax-related phenomenon. There have been few tests of intertemporal asset pricing theory. Those based on differences between equity and bond returns such as Hansen and Singleton(1982,1983) and Mehra and Prescott(1985) sharply rejected the theory while Marsh(1985) obtained more encouraging results in bond market data alone. The empirical evidence on the **APT** has been both generally supportive of and inconclusive regarding the implications of the theory. This will be discussed further in Section IV.

that the *unobservable* market portfolio is mean-variance efficient, a prediction with little pertinence for the behavior of the equally-weighted and value-weighted **CRSP** indices.

The Arbitrage Pricing Theory (**APT**) developed by Ross(1976,1977) represents one of the major attempts to overcome the problems with testability and the anomalous empirical that have plagued other theories. The main assumption of the theory is that returns can be decomposed into diversifiable and nondiversifiable components and that systematic risk can be measured as exposure to a small number of common factors. These strong *a priori* restrictions on the distribution of returns lead to an approximate theory of expected returns through the observation that capital market equilibrium should be characterized by the absence of riskless arbitrage opportunities.[2] Moreover, the **APT** can be applied to large subsets of the universe of risky assets which largely mitigates the problems raised by Roll(1977) regarding the testability of the **CAPM**.[3]

The apparent simplicity of the **APT** conceals serious difficulties associated with its empirical implementation. In particular, the theory cannot be tested on a given subset of available security returns without a strategy for measuring the common factors that are presumed to underlie security returns. In the absence of a theoretical specification of the components of systematic risk that can be related to observable economic data, most investigators have turned to the statistical method of factor analysis in order to implicitly

[2] There is considerable intellectual dissension concerning the extent to which the **APT** and its assumption of a linear factor model for security returns differs substantively from the one-period **CAPM** or intertemporal asset pricing models such as the one developed in Merton(1973). Pfleiderer(1983) provides a useful discussion of the nature of the differences between equilibrium based models such as the **CAPM** and the more statistically-based **APT**. Not surprisingly, given sufficient assumptions the predictions of these theories *can* intersect. For instance, if all asset returns follow a factor structure and there is a well-diversified portfolio on the efficient frontier, then both the **APT** and the **CAPM** will be true and the postulated factor model can aid in the measurement of the *unobservable* market portfolio. Similarly, continuous-time intertemporal asset pricing models can yield instantaneously linear factor pricing models that are observationally equivalent to the **APT**. Of course, the **APT** can be valid in settings where these alternative asset pricing models are false—that is the theories *need* not intersect.

[3] This observation does not remove *all* empirical ambiguities of the type discussed in Roll(1977). In particular, Shanken(1982,1985) has emphasized that the absence of riskless arbitrage opportunities coupled with the linear factor model for security returns places insufficiently exact restrictions on expected returns to lead to valid tests of the theory. Dybvig and Ross(1983) responded to this aspect of Shanken's criticism by noting that the additional assumptions needed to sharpen the testable implications of the theory are both mild and plentiful. Some of these assumptions are discussed in Section II.

measure these *unobservable* common factors. This solution circumvents the problems connected with the *a priori* stipulation of the sources of systematic risk by exchanging them for severe computational problems since it is prohibitively expensive to perform maximum likelihood factor analysis on large cross-sections. As a consequence, previous research has taken one of two courses: (*i*) performing maximum likelihood factor analysis on subgroups of thirty to sixty securities and then testing implications of the theory within and across subgroups and (*ii*) employing a less efficient statistical procedure such as principal components or instrumental variables to estimate the linear factor model in order to test the **APT** in large cross-sections. Since these procedures are not likely to yield portfolios which mimic the common factors as well as those produced by maximum likelihood factor analysis, this resolution of the problem of common factor measurement can lead to tests of the **APT** that reject the theory when it is, in fact, true.

This would seem to be a moot point since most empirical investigations have generally failed to reject various testable implications of the **APT**. Unfortunately, this observation is not cause for optimism since previous tests of the theory have suffered from two main defects. First, the practice of splitting the available collection of securities into small subgroups leads to weak tests of the **APT** both within and across subgroups. Second, all previous research has avoided confronting the **APT** with the full set of empirical regularities which have proved inexplicable by other theories. It seems fair to conclude that the current state of empirical knowledge regarding the **APT** is in an unsettled state.

In this paper, we propose to remove some of the empirical ambiguity surrounding the **APT** by performing comprehensive powerful tests of its implications. We can transcend some of the limitations of previous analyses through our ability to perform maximum likelihood factor analysis in large cross-sections, thus avoiding the need to split the universe of securities into subgroups or to resort to statistically inefficient estimation procedures. Moreover, our tests are constructed to be more powerful than those employed in previous research, both in their statistical formulation and in the choice of empirical regularities with which to confront the theory. As a consequence, we think that our work will provide incisive commentary on the validity of the **APT**.

The accomplishment of this ambitious task requires careful consideration of the theory and its implications as well as the explication of a cogent strategy for its implementation and testing. The next section provides a brief review of the **APT**. The third section describes our approach to maximum likelihood factor analysis as well as our procedure for forming portfolios which, in principle, constitute measurements of the components of systematic risk. The fourth section contains a detailed description of our tests for the validity of the theory and its diverse implications as well as a characterization of the difficulties associated with the determination of the number of common factors underlying security returns. This discussion is marked by systematic consideration of the power of alternative test procedures. The fifth section presents our empirical results while the final section is devoted to concluding remarks.

Some caveats are in order concerning the intended scope of this study. This paper provides a detailed examination of the validity of the **APT** and systematically ignores a number of interesting issues associated with its empirical implementation. We have addressed a number of these questions in other papers. In Lehmann and Modest(1985a), we compared the efficacy of a number of strategies for forming portfolios to mimic the factors postulated by the **APT** and determined that the method employed here performed best. Further evidence on this point may be found in Lehmann and Modest(1985b) which applied these same strategies as well as the usual **CAPM** benchmarks to the measurement of abnormal performance by mutual funds. That research verified that the statistical differences in the performance of the alternative portfolio formation strategies considered in Lehmann and Modest(1985a) translated into economically significant discrepancies in measured mutual fund performance. Lehmann and Modest(1985c) analyzes the appropriate frequency of observation for estimating factor models in order to construct portfolios to mimic the common factors over weekly and monthly intervals. There are other interesting aspects of the theory that have not yet been adequately dealt with. These questions include the predictive power and stationarity of the factor model of systematic risk and the ability of the common factors from one asset market to account for expected returns in other asset markets. Also much work needs to be done to link the unobservable common factors to observable economic data.[4] The investigation of these issues is on our research agenda.

[4] See Chen, Roll and Ross(1984) for one such preliminary investigation.

II. The Arbitrage Pricing Theory

The arbitrage theory of capital asset pricing developed by Ross (1976,1977), has aroused considerable interest in both the academic and business communities as a practical and testable alternative to the **Capital Asset Pricing Model**. In those papers, Ross persuasively argued that the key intuition underlying the **CAPM** was not the preference based analysis of Sharpe(1964), Lintner(1965), and Mossin(1966), but rather was the distinction between systematic and unsystematic risk inherent in the single index market model, introduced by Markowitz(1952) and developed and extended by Sharpe(1963,1967). Nowhere is this interpretation of the theory more clearly manifested than in the early empirical work on the **CAPM** performed by Fama and his students [see, in particular, Blume(1970) and Jensen(1969)]. In that literature it was conventional practice to justify the use of a proxy for the unobservable market by appealing to the ability of a well-diversified portfolio to mimic the market with negligible error when the market model provides an adequate description of security returns. As a consequence, tests of the **CAPM** using proxies for the market portfolio such as the **CRSP** equally-weighted index were interpreted as joint tests of the asset pricing theory and of the ability of the one factor model to characterize security returns.

Ross noted that there was no particular economic justification for the presumption that systematic risk can be adequately represented by a single common factor such as the return on the market. Instead, he assumed that systematic risk can be aggregated into **K** common factors and studied the implications of this assumption for expected returns. Hence, the distributional basis of the **APT** is that security returns are generated by the linear **K** factor model:

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

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

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, d_i, and is sufficiently independent across securities for a law of large numbers to apply.

The theory of asset pricing that naturally arises from the assumed return generating process follows from three key aspects of this formulation: (i) the linear relationship between individual security returns and factor and idiosyncratic risk; (ii) the number of securities whose returns follow this linear factor model is large (tending toward infinity); and (iii) the number of factors **K** is much smaller than the number of assets satisfying equation (1). The first point permits the decomposition of the risk of both individual securities and portfolios into the sum of systematic and idiosyncratic risk components. The second consideration suggests that well-diversified portfolios (i.e. those with weights of order $1/N$) will contain negligible idiosyncratic risk.[5] Finally, when the number of securities greatly exceeds the number of factors, it is easy to form well diversified portfolios which have no factor risk as well.

How do these features translate into an asset pricing theory? It follows from the observations made above that there are many (in the limit infinitely many) portfolios which have trivial (in the limit no) total risk so long as there are no taxes, transactions costs, or restrictions on short sales. Consequently, there will also be many ***zero net investment portfolios*** that have negligible total risk. As long as investors prefer more to less, these portfolios should earn zero profits to preclude riskless arbitrage opportunities. Since the number of arbitrage portfolios that can be formed grows without bound as the the number of securities satisfying the factor model (1) tends towards infinity, Ross and many others proved that the absence of riskless arbitrage opportunities implies that expected returns

[5] Note that this conclusion rests on the linearity of the return generating process as well since diversification need not eliminate idiosyncratic risk when returns are, for example, nonlinear in idiosyncratic risk.

must satisfy (*approximately*):

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

where:

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

$\lambda_k \equiv$ the risk premium on the k^{th} common factor,[6] $k = 1, \ldots, K$.

How are we to interpret the pricing intercept λ_0? Like the **CAPM**, the **APT** has a zero beta and a riskless rate formulation. However, unlike the **CAPM**, the difference between the two depends not on the availability of riskless borrowing and lending but on whether or not it is possible to form portfolios that are riskless from the countably infinite subset of risky assets under consideration. If it is possible to construct a portfolio that costs a dollar and has zero total risk then the intercept λ_0 corresponds to the riskless rate. The only way it will not be possible to form such a portfolio is if, under an appropriate normalization of the factor space, the factor loadings of all securities on one of the factors (the zero beta factor) are exactly the same. This would occur, for instance, if all security returns are equally affected by unexpected changes in a macroeconomic variable such as inflation or *GNP*.[7] In this case, λ_0 should be zero since the zero beta return is implicit in the linear factor model for security returns. In what follows, we will consider both the riskless rate and zero beta formulations in our empirical tests.

It is clear that the pricing relation (2) should price most assets with negligible error but need not price all assets arbitrarily well. If the pricing errors for *most* assets were not trifling, it would be easy to construct zero net investment arbitrage portfolios which were riskless and earned nonzero profits. Unfortunately, the same argument cannot be used to guarantee that all assets will be priced correctly since zero net investment portfolios must place appreciable weight on a *small* number of assets to exploit a *few* significant pricing deviations. Consequently, these portfolios will not be well-diversified and need not have negligible total risk. Similarly, these heuristic arguments can fail when applied to a large but finite number of assets since the constructed arbitrage portfolios will not be entirely

[6] Note that, if the return on a market index were the single common factor, then λ_1 would be the excess return on the market above λ_0.

[7] Formally this condition occurs when one of the eigenvectors of the covariance matrix of security returns contains identical elements.

riskless and very risk averse investors may not take advantage of nearly riskless arbitrage opportunities.

Not surprisingly, many investigators have examined the circumstances in which the pricing errors for all assets under consideration are negligible. Chamberlain and Rothschild(1983) proved that exact equality will obtain in an infinite economy setting if and only if there is a well-diversified portfolio on the mean-variance efficient frontier based on the (countably infinite) subset of returns which are presumed to satisfy the linear factor model. Connor(1984) and Shanken(1983) provide examples of equilibrium settings in which this occurs. Grinblatt and Titman(1983), Chen and Ingersoll(1983), and Dybvig(1983) provide explicit assumptions under which the equilibrium pricing errors can be computed in a finite economy setting. The results in these papers suggest the equilibrium pricing deviations will be small when the covariance between the marginal utility of wealth (or the derived marginal utility of wealth in an intertemporal asset pricing model) and residual risk is negligible.[8] This condition occurs if investors are not too risk averse and if the idiosyncratic risk of the individual assets and the value of each asset as a proportion of total wealth are not too large. In what follows, we assume sufficient structure to ensure that expected returns on the subset of risky securities we study (listed stocks on the New York and American Stock Exchanges) exactly satisfy the expected return condition (2).

III. Maximum Likelihood Factor Analysis and Basis Portfolio Formation

In this section we describe our approach to the estimation of the factor loadings, b_{ik}, and the idiosyncratic variances, d_i. We also detail our procedure for constructing basis portfolios from these estimates which are, in principle, highly correlated with the common factors that are presumed to be the dominant source of covariation among security returns. Unfortunately, there is a bewildering variety of estimation methods and portfolio formation procedures that have been advocated and used in the literature. In order to sort through the myriad of possibilities, we compared the efficacy of different combinations of basis

[8] The requirement that idiosyncratic risk be uncorrelated with investors' marginal utility of wealth is central to all utility-based asset pricing theories. In the **CAPM** framework, for instance, the assumption that asset returns follow a multivariate normal distribution or that investors have quadratic utility is the basis of the uncorrelatedness condition. In an intertemporal asset pricing context, similar assumptions lead to a lack of correlation between idiosyncratic risk and the derived utility of wealth.

portfolio formation procedures and estimation methods in Lehmann and Modest(1985a). We begin with a brief report on the outcome of that investigation.

In Lehmann and Modest(1985a), we sought to provide a comprehensive examination of different basis portfolio formation strategies. The estimation methods that we considered include maximum likelihood factor analysis, restricted maximum likelihood factor analysis,[9] principal components, and instrumental variables. We also examined four basis portfolio formation procedures: two variants of the Fama-MacBeth procedure and two versions of a quadratic programming method. Not surprisingly, we found that the maximum likelihood estimation procedures outperformed the less efficient instrumental variables and principal components methods. We also found that a simple variant of the Fama-MacBeth procedure provided performance at least as good as the more complicated quadratic programming methods and dominated the conventional Fama-MacBeth strategy. As a consequence, we will confine our attention to maximum likelihood factor analysis and the variant of the Fama-MacBeth procedure that produce what we refer to as minimum idiosyncratic risk portfolios.

The basis of maximum likelihood factor analysis is an assumption about the joint distribution of the factors and the security returns.[10] Given the **K** factor linear return generating process in (1), we can compactly write the demeaned returns of the N securities in matrix form as:

$$\underline{\tilde{r}}_t = \underline{\tilde{R}}_t - \underline{E} = B\underline{\tilde{\delta}}_t + \underline{\tilde{\epsilon}}_t \tag{3}$$

where $\underline{\tilde{r}}_t$ and $\underline{\tilde{R}}_t$ are $N \times 1$ vectors of security returns, $\underline{E}$ and $\underline{\tilde{\epsilon}}_t$ are $N \times 1$ vectors of expected security returns and residual risk respectively, B is an $N \times K$ matrix of factor loadings, and $\underline{\tilde{\delta}}_t$ is a $K \times 1$ vector of the time t realizations of the common factors. Under the assumption of joint normality of the returns $\underline{\tilde{r}}_t$ and the factors $\underline{\tilde{\delta}}_t$, the sample covariance matrix :

$$S = \frac{1}{T}\sum_{t=1}^{T} \underline{\tilde{r}}_t \underline{\tilde{r}}_t' \tag{4}$$

[9] By restricted maximum likelihood factor analysis, we mean maximum likelihood estimation of the factor analysis model for security returns subject to the restriction that expected returns are spanned by the factor loadings. The method is analogous to the maximum likelihood estimation of the zero beta **CAPM** employed, for example, by Gibbons(1982) and Stambaugh(1982).

[10] The basic reference on maximum likelihood factor analysis is Lawley and Maxwell (1971).

follows a Wishart distribution which serves as the basis of the log likelihood function:

$$\mathcal{L}(\Sigma|S) = \frac{-NT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{T}{2}\mathbf{trace}(S\Sigma^{-1}) \tag{5}$$

where:

$$\Sigma = \mathbf{E}[\tilde{r}_t\tilde{r}_t'] = BB' + D \tag{6}$$

under the usual assumptions of the statistical factor analysis model.[11] Maximum likelihood estimates of the factor loadings and idiosyncratic variances can be obtained by setting the derivatives equal to zero:

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial B} &= -T\Sigma^{-1}[\Sigma - S]\Sigma^{-1}B = 0 \\ \frac{\partial\mathcal{L}}{\partial D} &= -T\mathrm{Diag}[\Sigma^{-1}(\Sigma - S)\Sigma^{-1}] = 0 \end{aligned} \tag{7}$$

where $\mathrm{Diag}[X]$ is a diagonal matrix formed from the diagonal elements of X.

Due to the large number of distinct parameters in B and D $[N(K+1) - K(K-1)/2]$, iterative procedures for solving (7) are prohibitively costly. Conventional factor analysis therefore proceeds using the results of Joreskog(1967) who noted that given an estimate of D, it is possible to solve analytically for the maximum likelihood estimate of B under the normalization that $B'D^{-1}B$ is diagonal (which constitute the necessary $K(K-1)/2$ identifying restrictions on B). Joreskog showed that given D, the maximum likelihood estimate of B is:

$$\hat{B} = D^{1/2}\Omega(\Theta - I)^{1/2} \tag{8}$$

where Θ is a K dimensional diagonal matrix with the K largest eigenvalues of the matrix S^* $[S^* = D^{-1/2}SD^{-1/2}]$ along the diagonal and Ω is an $N \times K$ matrix of the corresponding eigenvectors.[12] Maximization of (5) then involves two steps : (i) given D, use the eigenvalue-eigenvector decomposition of S^* to arrive at new estimates of B and (ii) given B, solve (6) for D using (7). On convergence, the estimates of B and D are the required maximum likelihood estimates.

[11] By this, we mean the assumption that the covariance matrix of the idiosyncratic disturbances is diagonal. It is not possible to proceed with maximum likelihood estimation of the factor analysis model without the imposition of some such constraint.

[12] Joreskog noted that maximum likelihood factor analysis reduces to principal components when the idiosyncratic variances are identical. Consequently, conditional on the current estimate of D, the idiosyncratic variances of S^* are all unity.

A major shortcoming of standard maximum likelihood factor analysis is that it requires the repeated computation of the eigenvalues and eigenvectors of the $N \times N$ matrix S^*—a computationally infeasible procedure when the number of securities N is large. As an alternative, we employ the **EM** algorithm due to Dempster, Laird, and Rubin(1977) which was applied to factor analysis in Rubin and Thayer(1982). This procedure has the desirable feature that it involves only simple least squares regression operations and that the largest non-diagonal matrix inversion required is for a $K \times K$ matrix. Consequently, the **EM** algorithm can be used to handle larger cross-sections of securities than has heretofore been possible.

The **EM** algorithm follows from two simple observations. If the factors were observable, maximum likelihood estimation of B and D could proceed by a simple multivariate regression of the demeaned returns $\underline{\tilde{r}}_t$ on the factors $\underline{\tilde{\delta}}_t$. If instead B and D were observed, the factors could be estimated by their conditional expectation given $\underline{\tilde{r}}_t$. The key insight is that since each step is a conditionally maximizing step, iterative repetition of these steps is guaranteed to increase the log likelihood function (5)—a fact proven in Dempster, Laird, and Rubin(1977). This fact leads to a simple proof that the algorithm is guaranteed to converge to a local maximum of the likelihood function.

The reason that this procedure works is that the log likelihood function (5) can be factored into two parts corresponding to the two step iterative procedure sketched above. This occurs because the the density of $\underline{\tilde{r}}_t$ is the expected value of the joint density of $\underline{\tilde{r}}_t$ and $\underline{\tilde{\delta}}_t$ which, in turn, is the expected value of the conditional distribution of $\underline{\tilde{\delta}}_t$ given $\underline{\tilde{r}}_t$:

$$\begin{aligned} \Pr(\underline{\tilde{r}}_t|B,D) &= \int \Pr(\underline{\tilde{r}}_t, \underline{\tilde{\delta}}_t|B,D)\, d\underline{\tilde{\delta}}_t \\ &= \int \Pr(\underline{\tilde{\delta}}_t|\underline{\tilde{r}}_t, B, D)\Pr(\underline{\tilde{r}}_t|B,D)\, d\underline{\tilde{r}}_t \end{aligned} \tag{9}$$

Under the assumption that $\underline{\tilde{\delta}}_t$ and $\underline{\tilde{r}}_t$ are jointly normally distributed, the conditional distribution of $\tilde{\underline{\delta}}_t$ is a function of the usual sufficient statistics:

$$
\begin{aligned}
\mathbf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\underline{r}}_t\tilde{\underline{\delta}}_t'|\tilde{\underline{r}}_t\right] &= \frac{1}{T}\sum_{t=1}^{T}\tilde{\underline{r}}_t\tilde{\underline{r}}_t'\Lambda' \\
&= S\Lambda' \\
\mathbf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\underline{\delta}}_t\tilde{\underline{\delta}}_t'|\tilde{\underline{r}}_t\right] &= [I+B'D^{-1}B]^{-1} + \frac{1}{T}\sum_{t=1}^{T}\Lambda\tilde{\underline{r}}_t\tilde{\underline{r}}_t'\Lambda' \\
&= [I+B'D^{-1}B]^{-1} + \Lambda S\Lambda'
\end{aligned}
\tag{10a}
$$

where $\Lambda = [I+B'D^{-1}B]^{-1}B'D^{-1}$ and the statistics given in (10a) follow from noting that:

$$
\begin{aligned}
\mathbf{E}\left[\tilde{\underline{\delta}}_t|\tilde{\underline{r}}_t\right] &= \mathbf{E}\left[\tilde{\underline{\delta}}_t\right] + \mathbf{Cov}(\tilde{\underline{\delta}}_t,\tilde{\underline{r}}_t')\left[\mathbf{Var}(\tilde{\underline{r}}_t)\right]^{-1}\tilde{\underline{r}}_t \\
&= B'\Sigma^{-1}\tilde{\underline{r}}_t \\
&= [I+B'D^{-1}B]^{-1}B'D^{-1}\tilde{\underline{r}}_t \\
\mathbf{Var}(\tilde{\underline{\delta}}_t|\tilde{\underline{r}}_t) &= \mathbf{Var}(\tilde{\underline{\delta}}_t) - \mathbf{Cov}(\tilde{\underline{\delta}}_t,\tilde{\underline{r}}_t')\left[\mathbf{Var}(\tilde{\underline{r}}_t)\right]^{-1}\mathbf{Cov}(\tilde{\underline{r}}_t,\tilde{\underline{\delta}}_t') \\
&= I - B'\Sigma^{-1}B \\
&= [I+B'D^{-1}B]^{-1} \\
\mathbf{E}[\tilde{\underline{\delta}}_t\tilde{\underline{\delta}}_t'|\tilde{\underline{r}}_t] &= \mathbf{Var}(\tilde{\underline{\delta}}_t|\tilde{\underline{r}}_t) + \mathbf{E}[\tilde{\underline{\delta}}_t|\tilde{\underline{r}}_t]\mathbf{E}[\tilde{\underline{\delta}}_t|\tilde{\underline{r}}_t]'
\end{aligned}
\tag{10b}
$$

Having computed the expected value of the likelihood function given B and D using the sufficient statistics (10a), it is straightforward to maximize it with respect to B and D to obtain:

$$
\begin{aligned}
\hat{B} &= \mathbf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\underline{r}}_t\tilde{\underline{\delta}}_t'|\tilde{\underline{r}}_t\right]\ \mathbf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\underline{\delta}}_t\tilde{\underline{\delta}}_t'|\tilde{\underline{r}}_t\right]^{-1} \\
\hat{D} &= \mathbf{Diag}\left[S - B\ \mathbf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\underline{\delta}}_t\tilde{\underline{\delta}}_t'|\tilde{\underline{r}}_t\right]B'\right]
\end{aligned}
\tag{11}
$$

Each iteration consists of of the evaluation of the expected value of the usual sufficient statistics for a multivariate linear regression utilizing (10) and employing those statistics to perform the regression as in (11). Note that the iteration between (10) and (11) only requires inversion of the $K \times K$ matrix $[I + B'D^{-1}B]$ and the diagonal matrix D. On convergence, the estimates of B and D are the required maximum likelihood estimates.[13]

[13] We define convergence as a stationary point such that the sum of the squared derivatives given in (7) is less than 0.0001. We use as our starting estimates of B and D instrumental variables estimates based on a procedure outlined in Lehmann and Modest(1985a).

The estimation of the factor loadings and idiosyncratic variances is only half of the usual two-step procedure for testing the **APT**. Following estimation, it is conventional practice to construct portfolios whose returns are, in principle, highly correlated with the factors. For example, most studies perform a series of cross-sectional regressions by weighted least squares at each date t of security returns on the estimated factor loadings using the estimated idiosyncratic variances as weights. The time series of coefficients from these regressions can be interpreted as the returns on portfolios which are highly correlated with the unobservable common factors. The procedure that we have found to work best is a variant of the usual Fama-MacBeth cross-sectional regression procedure. In the jargon of optimal portfolio formation, our method, which produces what we call minimum idiosyncratic risk portfolios, involves choosing the N portfolio weights $\underline{w}_j$ to mimic the j^{th} factor so that they:

$$\min_{\underline{w}_j} \underline{w}_j{}' D \underline{w}_j \tag{12}$$

subject to:

$$\begin{aligned} \underline{w}_j' \underline{b}_k &= 0 \qquad \forall j \neq k \\ \underline{w}_j' \underline{\iota} &= 1 \end{aligned} \tag{13}$$

where $\underline{b}_k$ is the k^{th} column of the factor loading matrix B, D is the diagonal matrix consisting of the variances of the idiosyncratic disturbances, and $\underline{\iota}$ is a vector of ones.[14]

How do these portfolios differ from the more familiar Fama-MacBeth portfolios? The answer lies in the scaling of the two portfolios. Fama-MacBeth portfolios are the sample minimum idiosyncratic risk portfolios which have a loading of *one* on one factor and loadings of zero on the other factors (prior to rescaling to unit net investment). Minimum idiosyncratic risk portfolios also are constructed to have sample loadings of zero on the same factors, but their only other requirement is that they cost a dollar. As a consequence, minimum idiosyncratic risk portfolios need not have any particular loading on the factor being mimicked.

In Lehmann and Modest(1985a) we demonstrated that if the population values of B and D are used to construct the basis portfolios then, under mild assumptions, the

[14] 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 $B^* = (\underline{b}_1 \underline{b}_2 \ldots \underline{\iota} \ldots \underline{b}_k)$, $\underline{\iota}$ is a vector of ones in the j^{th} column, and $\underline{e}_j$ is a vector of zeros except for a one in the j^{th} position.

Fama-MacBeth reference portfolios will do a better job of mimicking the factors than the minimum idiosyncratic risk portfolios in that the Fama-MacBeth portfolios will be more highly correlated with the true unobservable factors. Of course, in actual practice we must substitute estimates of B and D for the corresponding population values to form basis portfolios. In this case, the Fama-MacBeth portfolios need not do a superior job of mimicking the factors and, in fact, the minimum idiosyncratic risk portfolios may be more highly correlated with the factors.

Why might this occur? The answer lies in the requirement that Fama-MacBeth portfolios have a sample loading of one on the factor being mimicked. The Fama-MacBeth procedure thus tends to place relatively large weight on securities with large factor loading estimates and less weight on those with small sample factor loadings. If there is little measurement error in the sample factor loadings, this procedure will yield good basis portfolios since the returns of securities with large factor loadings are more informative about fluctuations in the underlying common factors. Unfortunately, in the presence of measurement error, large factor loading estimates can reflect large measurement error while small factor loading estimates can occur when measurement error offsets otherwise large true factor loadings. Hence, the Fama-MacBeth procedure need not place the appropriate weights on individual securities in the presence of measurement error. In contradistinction, the minimum idiosyncratic risk procedure is unaffected by the presence of measurement error in the factor loadings.[15] The comprehensive evidence presented in Lehmann and Modest(1985a) suggests that measurement error in the loadings is sufficiently pernicious to warrant employment of the minimum idiosyncratic risk procedure.[16]

[15] It might appear that both methods would be quite sensitive to measurement error in the factor loadings due to their common requirement that the portfolio weights be orthogonal to the sample loadings of the other factors [i.e. $\underline{w}_j'\underline{b}_k = 0 \quad \forall j \neq k$]. Fortunately, this condition imposes no real constraint; rather it merely helps to determine a particular *sample* rotation or normalization of the factor estimates. Both procedures can be sensitive to measurement error in the idiosyncratic variances. We have ignored this problem because we surmise that the application of weighted least squares when the weights are measured with error will still typically yield good basis portfolios, an intuition that is based on conventional econometric wisdom surrounding heteroskedastic regression models.

[16] One other problem with the Fama-MacBeth procedure is worth noting. In factor model estimation, it is conventional practice to normalize the factors so that they are uncorrelated and have unit variances and to normalize the factor loadings so that $B'D^{-1}B$ is diagonal. This practice yields typical factor loading estimates that are much less than one–on the order of .001 to .0001 in daily data. As a consequence, to ensure that $\underline{w}_j'\underline{b}_j = 1$ and

Finally, the implementation of the riskless rate version of the **APT** requires the measurement of portfolios that are orthogonal to the common factors. Such portfolios should earn the riskless rate if this version of the **APT** is true. Not surprisingly, we will choose the N portfolio weights $\underline{w}_{rf}$ to mimic the approximately riskless portfolio of risky assets so that they:

$$\min_{\underline{w}_{rf}} \underline{w}_{rf}{}' D \underline{w}_{rf} \tag{14}$$

subject to:

$$\begin{aligned} \underline{w}'_{rf}\underline{b}_k &= 0 \qquad \forall k \\ \underline{w}'_{rf}\underline{\iota} &= 1 \end{aligned} \tag{15}$$

where, again, $\underline{b}_k$ is the k^{th} column of the factor loading matrix B, D is the diagonal matrix consisting of the variances of the idiosyncratic disturbances, and $\underline{\iota}$ is a vector of ones. This is precisely the portfolio for the intercept that is produced by the Fama-MacBeth style cross-sectional regression on a constant and the factor loadings.

IV. Hypothesis Testing Procedures

A. Previous Tests of the APT

Before outlining our procedures for testing the **APT**, it is certainly worth reflecting on the procedures used in previous studies to assess the validity of the theory. The main purpose of this review is to obtain some guidance regarding the power of different testing strategies. As a consequence, we simply sketch some of the procedures employed by others and suggest reference to the original sources for more detailed discussion.

Since the **APT** only requires that security returns satisfy an approximate rather than an exact factor structure, all of the testable implications of the theory (in a finite sample of securities) lie in the restriction given by equation (2) that expected returns are spanned

---

$\underline{w}'_j\underline{b}_k = 0 \quad \forall j \neq k$ the Fama-MacBeth procedure must place large positive and negative weights on at least some securities. For instance, we have found that the Fama-MacBeth procedure frequently produces portfolio weights in excess of one hundred percent in absolute value. Thus straightforward application of the Fama-MacBeth strategy under the conventional normalization of the factor model can yield poorly diversified portfolios. Clearly this problem can amplify the impact of even otherwise trivial measurement error in the factor loadings. In Lehmann and Modest(1985a), we provide an alternative normalization which mitigates this particular problem. Unfortunately, it does not resolve the difficulties discussed in the text.

by the factor loadings. Previous tests of this mean restriction have typically taken three forms: (i) tests for the equality of intercepts across small subgroups of securities, (ii) tests for the joint significance of the factor risk premia in each subgroup and (iii) tests for the insignificance of nonfactor risk measures in explaining expected returns. The first two types of tests involve cross-sectional weighted least squares regressions of the returns of the securities in each subgroup on a constant and the estimated factor loadings using the estimated idiosyncratic variances as weights. If the **APT** is true, then the time series means of the intercepts from these regressions should, apart from sampling error, be identical across groups and equal to the riskless rate if the riskless rate version of the theory is correct or equal to zero when the zero beta formulation is appropriate. In addition, the factor risk premia in each group should be jointly significantly different from zero in large samples.[17] The third type of test usually takes the form of a similar cross-sectional regression in which the returns in each subgroup are regressed on the factor loading estimates and either the estimated idiosyncratic or total standard deviation of individual security returns. If the **APT** is true, then the time series mean of the coefficients on the nonfactor risk term should be insignificantly different from zero.

Most of the existing empirical literature on the validity of the **APT** has failed to provide substantive evidence against the theory. Unfortunately, this body of work suffers in large part from a serious problem: the tests often lack the power to reject the theory when it is false. Some of the problems with earlier tests are a consequence of the technological necessity of dividing the universe of securities into small groups to perform maximum likelihood factor analysis with conventional software packages.[18] This forced reliance on small cross-sections has two deleterious consequences. First, it results in imprecise estimates of the pricing intercepts λ_0 and the factor risk premia $\underline{\lambda}$ that render statistical tests involving these quantities particularly susceptible to Type II errors. Second, the reliance

[17] Since the factors are not observable, most studies begin by performing factor analysis on each subgroup of securities. Unfortunately, the sample rotation of the factors may not be the same across different factor analysis runs. Consequently,there is no prediction that the factor risk premia should be equal across groups and the only testable restriction across groups is whether the intercepts are equal.

[18] This is especially true of most of the studies which have performed factor analysis on numerous subgroups of thirty to sixty securities, including Roll and Ross(1980), Hughes(1982), Brown and Weinstein(1983), Dhrymes, Friend, and Gultekin(1984), and Dhrymes,Friend, Gultekin, and Gultekin(1985).

on small cross-sections prevents the implementation of tests that have proven powerful in the **CAPM** context—such as the examination of the risk-adjusted returns on portfolios sorted on the basis of some characteristic such as dividend yield, price-earnings ratio, or size.[19] Finally, there are more substantive problems reflecting the difficulty in constructing powerful mean-variance efficiency tests given the large variability in equity returns.

In the first comprehensive empirical investigation of the **APT**, Roll and Ross(1980) performed all three types of tests. Like many subsequent authors, they tested for the equality of the intercepts across subgroups and failed to reject the null hypothesis of equality. As noted by Roll and Ross(1980), however, these tests have little power since the sampling variation in the estimated intercepts is so large that it would be difficult to reject almost any reasonable hypothesis about them. In their test for the significance of at least one of the factor risk premia, Roll and Ross found 88.1% of the portfolios had at least one significant factor risk premium at the 5% significance level.[20] This, however, is not really a test of the **APT** in that it is an implication that is consistent with many other asset pricing theories as well. Roll and Ross also performed the third type of test and examined the impact of unsystematic risk (represented by own variance) on the pricing of assets in addition to the effect of systematic risk exposure captured by the **APT** factor loadings. In this test, they also failed to reject the null hypothesis. Once again the problem is that we know this test has little power. Fama and MacBeth(1973) studied whether idiosyncratic variance had an additional impact on expected returns over the explanatory power of beta in their **CAPM** tests. If the multi-factor **APT** is true as postulated by Roll and Ross(1980), Fama and MacBeth(1973) should have rejected the adequacy of the single index market model since the estimated idiosyncratic variances should have reflected, in part, the loadings on the omitted factors. Of course, Fama and MacBeth(1973) failed to reject the mean-variance efficiency of the equally weighted index using just such a test.

Chen(1983) avoided the problems associated with splitting the universe of securities into subgroups by employing an inexpensive instrumental variables estimator to obtain es-

[19] See, for instance, Cannistraro(1973), Basu(1977), Litzenberger and Ramaswamy(1979), Banz(1981), and Reinganum(1981a).

[20] These findings have been challenged by Dhrymes, Friend, Gultekin and Gultekin(1985) who, using basically the same techniques and data set, found only three of forty-two portfolios with significant premia.

timates of the factor loadings.[21] His testing procedure was quite simple: divide the universe of securities into two groups after ranking on some characteristic and form portfolios from these groups with identical estimated factor loadings. If the **APT** is true, these pairs of portfolios should have identical mean returns. Chen rejected the hypothesis that the factor loading adjusted portfolios have identical mean returns in one of the four (four year) subperiods based on both firm size and previous period return and found no rejections based on own variance. These results were interpreted to be largely supportive of the **APT**. The problem with this test is again one of power. Even when no adjustment was made for risk, the differences in the mean returns on equally weighted portfolios from both the high and low firm size and own variance groups are statistically significant in only two of four subperiods.[22] The insignificance of the mean return differences of these pairs of portfolios stands in sharp contrast to the corresponding differences we found in mean returns of portfolios constructed from the first and fifth size or variance quintiles. The mean return differences based on these extreme quintiles are statistically significant in all subperiods perhaps due to the nonlinearity of the own variance or size effects. Chen was aware of this potential problem and reported that the results were similar when the significance tests were based on the top and bottom deciles of firm size and own variance.[23] Nevertheless, the possible difficulties suggest considerable caution in the choice of a testing strategy.

B. Testing the APT

We employ conventional multivariate test statistics to test the **APT**. For reasons outlined above, care must be taken to insure that the tests have adequate power. Our maximum likelihood factor analysis procedure permits us to estimate factor models for large cross-sections of security returns. As a consequence, we can obtain efficient estimates of factor loadings and idiosyncratic variances without suffering from the technological necessity of splitting the cross-section into thirty to sixty security subgroups or resorting to inefficient

[21] These estimates are less efficient than those obtained from maximum likelihood factor analysis of the same cross-section but Chen presumed that the loss in efficiency was small relative to the gain associated with working with substantially larger cross-sections than had previously been possible.

[22] The comparable information is not available for the previous period return results.

[23] However, this claim stands in sharp contrast to our results discussed below where the conclusions are quite sensitive to whether five, ten, or twenty portfolios are used. In addition, we find significantly different risk adjusted returns between the top and bottom firm size deciles in each of our five year periods.

estimation procedures. This ability to perform maximum likelihood factor on large subsets of securities then allows to perform relatively efficient tests of the **APT** mean restriction using sorted portfolios based on characteristics that have proven to be effective in analyzing the mean-variance efficiency of the standard **CAPM** benchmarks and to avoid the problems associated with testing for the equality of intercepts or the significance of factor risk premia across small subgroups that have plagued previous studies. To guard against the potential nonlinearity of the dividend yield, own variance and size effects and any possible adverse power consequences, we group securities into different numbers of portfolios based on these characteristics. In addition, we provide further information concerning the power of our tests by using similar procedures to test the mean-variance efficiency of the equally-weighted and value-weighted **CRSP** indices. For example, the failure to reject the **APT** and simultaneous rejection of the mean-variance efficiency of the usual market proxies would suggest that our tests have good power against reasonable alternatives and should be taken as serious evidence in favor of the **APT**.

We implement the tests in the following manner. First, we estimate a factor model for security returns using the method of maximum likelihood and then employ the minimum idiosyncratic risk procedure to form basis portfolios. Second, we form portfolios of securities ranked on characteristics such as firm size, dividend yield, and own variance. We consistently formed five, ten, and twenty such portfolios for testing purposes to guard against potential power difficulties although we do not report the ten portfolio results to conserve space. We then run the regression of raw or excess portfolio returns on the corresponding basis portfolio returns and a constant. The usual F test for the hypothesis that the intercepts for each portfolio are jointly insignificantly different from zero provides a test of either the riskless rate or zero beta version of the **APT**.

More formally, let $\underline{\tilde{R}}_{pt}$ be the vector of excess returns on the sorted characteristics portfolios when the riskless rate version of the **APT** is true and be the corresponding raw returns when the zero beta version is appropriate. Similarly, let $\underline{\tilde{R}}_{mt}$ be the corresponding returns on the basis portfolios. Consider the fitted multivariate regression of $\underline{\tilde{R}}_{pt}$ on $\underline{\tilde{R}}_{mt}$ and a constant term:

$$\underline{\tilde{R}}_{pt} = \underline{\hat{\alpha}}_p + \hat{B}_p \underline{\tilde{R}}_{mt} + \underline{\hat{\epsilon}}_{pt} \tag{16}$$

where $\underline{\hat{\alpha}}_p$ is the estimated constant term vector, $\hat{B}_p$ is the estimated factor loading matrix, and $\hat{\underline{\epsilon}}_{pt}$ is the fitted residual vector. If the **APT** is true and the basis portfolios are measured without error, then $\hat{\underline{\alpha}}_p$ should be statistically insignificantly different from zero. On the assumption that $\tilde{\underline{R}}_{mt}$ and $\tilde{\underline{R}}_{pt}$ are both normally distributed random vectors, the usual F statistic for testing this hypothesis is:

$$\frac{\hat{\underline{\alpha}}_p' \hat{\Omega}_p^{-1} \hat{\underline{\alpha}}_p}{1 + \overline{\underline{R}}_m' \hat{\Sigma}_m^{-1} \overline{\underline{R}}_m} \left[\frac{T(T-K-N)}{N(T-K-1)} \right] \qquad \sim F(N, T-K-N) \tag{17}$$

where $\hat{\Omega}_p$ is the sample residual covariance matrix of $\hat{\underline{\epsilon}}_{pt}$, $\overline{\underline{R}}_m$ is the vector of sample mean returns on the basis portfolios, and $\hat{\Sigma}_m$ is the sample covariance matrix of their returns. This is our basic **APT** test—all variations we perform provide informal checks on the power of the tests.[24]

Of course, a critical assumption is that the basis portfolios are measured without error, an assumption that is only correct as the number of securities in the first stage factor analysis tends toward infinity. Fortunately, the large number of securities in our cross-sections suggests that the measurement error in our basis portfolios is likely to be small.[25] Moreover, any measurement error causes these F statistics to be biased toward rejection when the **APT** is true, although the magnitude of this bias is likely to be trivial relative to the measurement error in $\hat{\underline{\alpha}}_p$.

It is worth mentioning an alternative strategy which partially mitigates the effects of

[24] The tests employed here for the mean-variance efficiency of the usual **CAPM** indices differ somewhat from (17). We employ Shanken's(1984) multivariate test statistic which is of the form of (17) with $K = 1$ and where N is replaced with $N - 1$. There is a more substantive difference in the computation of $\hat{\underline{\alpha}}_p$. For the mean-variance efficiency tests, $\hat{\underline{\alpha}}_p$ is the vector of residuals from the cross-sectional regression of the portfolio market model intercepts on one minus their estimated betas. See the discussion below for an indication of a possible reduction in the power of the tests associated with the estimation of the zero beta rate from the *sorted* portfolios.

[25] It is worth noting that we can provide large sample estimates of the measurement error covariance matrix of our basis portfolios in order to get a feel for the likely severity of this problem. From the solution to equation (14), the idiosyncratic risk component in the j^{th} basis portfolio has the approximate variance: $\hat{\sigma}_j^2 \approx \underline{e}_j'(\hat{B}^{\star\prime} \hat{D}^{-1} \hat{B}^\star)^{-1} \underline{e}_j$ where $\hat{B}^\star$ is the matrix with the estimated factor loadings in each column except for the j^{th} column, which is a column of ones, $\hat{D}^{-1}$ is the diagonal matrix of estimated idiosyncratic variances, and the approximation arises because we are using estimates instead of the true values of the relevant parameters. We have examined these approximate variances and the associated covariances and have found that these numbers are uniformly trivial across both time periods and factor models. Hence, it seems reasonable to conclude that measurement error is not a particularly serious problem in our large cross-sections.

measurement error that we have chosen to ignore due to its deleterious impact on the power of our tests. We could estimate the basis portfolio returns by straightforward application of the Fama-MacBeth style generalized least squares cross-sectional regression of the *sorted* portfolios' returns on a constant and their estimated factor loadings from the first stage factor analysis at each date t. The analogue to $\underline{\hat{\alpha}}_p$ can then be computed from the time series mean of the residuals. This would tend to alleviate the measurement error problem discussed above since the measurement error in the factor loadings of the sorted portfolios should be much smaller than the typical error in the loadings of the individual securities, given that the portfolios are well diversified and not formed on the basis of their sample loadings. Following the analysis in Shanken(1984), we could then construct the large sample F statistic just as in (17) above with two main modifications: (i) the sample mean vector and covariance matrix of the cross-sectional regression coefficients replace those of $\underline{\bar{R}}_{mt}$ and (ii) the degrees of freedom in the numerator of the F statistic are reduced by $K+1$ due to the estimation of the factor risk premia.[26]

Why do we eschew this seemingly superior statistical procedure? The reason is once again a matter of power. The problem with this procedure is that it involves estimation of the factor risk premia using the portfolios formed from well-known empirical anomalies. Suppose that the **APT** is false and we construct the test statistics in this revised fashion. The generalized least squares cross-sectional regressions will choose estimates of the factor risk premia which minimize the weighted sum of squared residuals. This, in turn, will tend to make the F statistic small and, hence, can cause a failure to reject the null hypothesis when it is false.[27] In our procedure, we estimate the factor risk premia from the whole sample of securities underlying the factor analysis, a sample that is not biased with regard to firm size, dividend yield, or own variance. These premia are then used to estimate $\underline{\hat{\alpha}}_p$ and to test its significance. If the **APT** is false, our procedure will not bias the estimates of $\underline{\hat{\alpha}}_p$ toward acceptance of the null hypothesis.

[26] The only modification to Shanken's analysis involves the fact that the factor loadings are not coefficients from a linear regression as in the case of market model estimates but rather are estimates of factor loadings from a nonlinear maximum likelihood procedure. We doubt that this would affect the analysis appreciably, especially in large samples.

[27] This reduction in power will not occur when the multiple correlation between the sorting characteristics and the factor loadings and a vector ones is zero. This cannot occur in our samples since our anomalies are non-negative and, hence, will at least be correlated with the intercept.

This would seem to be a trivial debating point except for the fact that we have encountered precisely this problem in testing the mean-variance efficiency of the usual market proxies. If, by analogy with Fama and MacBeth(1973), we estimate the zero beta **CAPM** by regressing portfolio returns on portfolio betas plus an intercept, we often fail to reject the mean-variance efficiency of either **CRSP** index using Shanken's multivariate test statistic. If we follow Black, Jensen, and Scholes(1972) and perform a cross-sectional regression of the portfolio intercepts from market model time series regressions on one minus their corresponding sample betas, we typically sharply reject mean-variance efficiency at extremely low marginal significance levels using the appropriate variant of Shanken's test statistic. Similarly, we would *never* reject the **APT** in our samples if we estimated the factor risk premia from cross-sectional regressions of the characteristics-based portfolios to construct the relevant test statistics. Consequently, we have chosen to ignore the impact of measurement error on our test statistics due to their greater power.

C. Comparing the Riskless and Zero Beta Versions of the APT

If the riskless rate formulation of the **APT** is true, then security returns satisfy:

$$\underline{\tilde{R}}_t - \underline{\iota} R_{ft} = B(\underline{\tilde{R}}_{mt} - \underline{\iota} R_{ft}) + \underline{\epsilon}_t \tag{18}$$

where R_{ft} is the return on the limiting riskless portfolio of risky assets whose mean return is λ_0. If the zero beta version of the **APT** is true, then security returns satisfy:

$$\underline{\tilde{R}}_t = B\underline{\tilde{R}}_{mt} + \underline{\epsilon}_t \tag{19}$$

since the zero beta portfolio corresponds to one of the common factors underlying security returns.

How can we distinguish the two versions of the **APT**? The simplest answer involves asking under what circumstances the two equations for returns are identical. Evidently, the riskless rate and zero beta models are the same when:

$$B\underline{\iota} = \underline{\iota} \tag{20}$$

that is, when the sum of the factor loadings for each security is one. This conclusion follows from simple efficient set reasoning as well: since the basis portfolios $\underline{\tilde{R}}_{mt}$ span the efficient set when the zero beta formulation is correct, the sum of the factor loadings must be one because, under appropriate rotation, the factor loading corresponding to a zero beta factor is $1-\beta$ and the remainder sum to β when appropriately weighted so that the remaining factors add up to an efficient portfolio.

As a consequence, a first test of the zero beta **APT** is to see whether the sum of the coefficients in either equation (18) or (19) is unity. This is difficult if we try to aggregate the test statistics from individual security regressions due to the presence of industry effects in the residual covariance matrix. While we could try to account for this problem in constructing the aggregate test statistic, it is simpler to perform the test on portfolios instead. There is a potential loss of power associated with moving from individual securities to portfolios. Fortunately, there is a gain in precision in estimating portfolio loadings as well as the sum of the portfolio loadings which should permit reasonably powerful tests. On the assumption that $\underline{\tilde{R}}_{mt}$ and $\underline{\tilde{R}}_{pt}$ are both normally distributed random vectors, the usual F statistic for testing this hypothesis is:

$$\frac{(\hat{B}_p \underline{\iota}_k - \underline{\iota}_n)' \hat{\Omega}_p^{-1} (\hat{B}_p \underline{\iota}_k - \underline{\iota}_n)}{\underline{\iota}_k' \hat{\Sigma}_m^{-1} \underline{\iota}_k} \left[\frac{T-K-N}{N}\right] \sim F(N, T-K-N) \tag{21}$$

where $\hat{B}_p$ is the matrix of estimated portfolio factor loadings, $\hat{\Omega}_p$ is the sample residual covariance matrix of the regression residuals, and $\underline{\iota}_k$ and $\underline{\iota}_n$ are vectors with K and N unit elements respectively. We perform this test on both the excess return and raw return regressions to guard against possible differences in the powers of the two test formulations.

This test does not exhaust the differing implications of the riskless rate and zero beta models. Each version of the **APT** places different restrictions on λ_0—the intercept in the pricing equation. The riskless rate formulation suggests that λ_0 is equal to the riskfree rate while the zero beta version implies that λ_0 is zero. In this spirit, we report summary statistics for the orthogonal portfolios with weights $\underline{w}_{rf}$ defined in (14) and (15) above. We simply test whether the mean return on this portfolio and whether the mean difference in return between this portfolio and the riskfree asset are significantly different from zero to compare the two models in this other dimension. Of course, previous examinations of returns on these orthogonal portfolios have typically failed to reject *both* hypotheses but we are confident that our large cross-sections will permit us to measure the mean returns on these orthogonal portfolios with considerable precision. This test also provides another test of the **APT**: *both* versions of the theory will be rejected if mean returns on the orthogonal portfolios are significantly larger than the riskless rate or significantly negative. The fact that estimated zero beta rates are typically significantly greater than the riskfree rate in **CAPM** studies such as Black, Jensen, and Scholes(1972), Fama and MacBeth(1973), Litzenberger and Ramaswamy(1979), Gibbons(1982), and Stambaugh(1982) suggests that this is not an unlikely possibility.

It is worth noting that measurement error has a greater effect on tests involving orthogonal portfolios than it does on the F tests for the **APT** itself. The problem is that the basis portfolios will tend to mimic *some* rotation of the factor space even if it is not the one that was assumed for the purposes of estimation. This occurs because idiosyncratic risk is likely to be virtually eliminated in well diversified portfolios of large numbers of securities while sensitivity to *all* of the common factors is likely to remain in each basis portfolio as long as the individual loadings contain some measurement error. In a similar vein, the orthogonal portfolios are constructed to have weights orthogonal to estimated factor loadings, but they are not necessarily orthogonal to the true loadings. Thus to the extent that measurement error is present in the factor loading estimates, some factor risk is likely to remain in these orthogonal portfolios as well.

The one factor case provides a convenient setting for analyzing the potential impact of measurement error. In this case, security returns satisfy:

$$\underline{\tilde{R}}_t = \underline{\iota} R_{ft} + \underline{\beta}(\tilde{R}_{mt} - R_{ft}) + \underline{\tilde{\epsilon}}_t \tag{22}$$

assuming the riskless rate version of the theory is appropriate. Unfortunately, we do not know the betas and instead must estimate them. Let the measured betas be unbiased estimates of the true betas so that:

$$\underline{b} = \underline{\beta} + \underline{v} \tag{23}$$

where the elements of the measurement error vector $\underline{v}$ have zero mean and finite variances and covariances. We assume that the factor loadings have been normalized so that $\underline{b}'\underline{b} = \underline{b}'\underline{\iota}$. For simplicity, we also assume that the sample mean of the measurement errors across beta estimates is zero [i.e. $\overline{v} = \frac{1}{N}\sum_{i=1}^{N} v_i \approx 0$].

What happens if we use the estimated betas in an attempt to mimic the approximately riskless portfolio of risky assets? If we assume for simplicity that the idiosyncratic variances are identical [i.e. $D = \sigma_\epsilon^2 I$], then the portfolio weights which solve (14) and (15) are:

$$\underline{w}_{rf} = \frac{1}{N(\bar{b}-1)}(\underline{b} - \underline{\iota}) \tag{24}$$

where $\bar{b}$ is the average sample beta. It is straightforward to verify that this portfolio contains factor risk since its true loading is:[28]

$$\beta_{rf} = \frac{\sigma_\beta^2 + \bar{b}(\bar{b}-1) + \frac{1}{N}\underline{v}'\underline{\beta}}{\bar{b}-1} \tag{25}$$

where $\sigma_\beta^2 = \frac{1}{N}\sum_{i=1}^{N}(\beta_i - \bar{\beta})^2$ is the sample variance of the true betas. Given the additional assumption that $\frac{1}{N}\underline{v}'\underline{\beta}$ is approximately zero (a reasonable assumption in large cross-sections), this approximately simplifies to:

$$\beta_{rf} \approx \bar{b} + \frac{\sigma_\beta^2}{\bar{b}-1} = \frac{\sigma_\beta^2 - \sigma_b^2}{\bar{b}-1} \tag{26}$$

which will be positive.[29] As a consequence, the mean return on this portfolio will be biased upward in periods when the sample market risk premium is positive and will be biased downward in periods when the average market risk premium is negative.

This conclusion is not surprising: it is well known in a regression setting that, if the independent variable is measured with error, the slope coefficient is biased downward and the intercept is biased in the direction of the sign of the mean of the independent variable. It is equally clear that, in this example, the magnitude of the bias can be quite serious, depending on how close $\bar{b}$ is to one (i.e. how close the dispersion in the elements b_i are to zero). Of course, the magnitude of the bias depends on additional considerations in a multifactor setting with nontrivial industry effects in the residual risks. Nevertheless, this example suggests that we should interpret any rejections of this formulation of the **APT** with considerable caution.

D. Determining the Number of Factors Underlying the APT

Surprisingly, the most difficult empirical problem in the **APT** is the determination of the number of factors underlying security returns. The problem is insidious: the test which

[28] This expression uses the assumption that $\bar{b} \approx \bar{\beta}$ made above ($\bar{v} \approx 0$).

[29] This expression is positive since $0 < \bar{b} < 1$ under the normalization $\underline{b}'\underline{b} = \underline{b}'\underline{\iota}$ to insure that $\sigma_b^2 > 0$, and $\sigma_b^2 > \sigma_\beta^2$ when $\frac{1}{N}\underline{v}'\underline{\beta} = 0$.

is powerful can reject the hypothesis that there are K pervasive sources of risk even when it is true while the remaining tests have little power to reject the null hypothesis when it is false. We will examine several possible tests for the number of factors and will catalogue the weaknesses of the various approaches.

The most obvious approach is to use the likelihood ratio test statistic for the number of factors that is widely employed in the psychometric literature. This test statistic was studied by Bartlett(1950) and was the subject of a Monte Carlo investigation by Geweke and Singleton(1980). The test is simple—under the assumption that the stationary ***statistical*** factor analysis model describes security returns (i.e. that there is no correlation among the idiosyncratic disturbances), the covariance matrix of security returns, Σ, can be written as in (6) $[\Sigma = BB' + D]$ where D is diagonal while under the alternative hypothesis Σ has no particular structure. The null hypothesis may be tested by minus twice the logarithm of the likelihood ratio for this hypothesis:

$$\chi^2(q) = T[\ln|\hat{B}\hat{B}' + \hat{D}| + \mathbf{trace}[S(\hat{B}\hat{B}' + \hat{D})^{-1}] - \ln|S| - N] \tag{27}$$

which, in large samples, is approximately distributed as χ^2 with q degrees of freedom where $q = [(N-K)^2 - (N+K)]/2$. Bartlett(1950) showed that the distribution of the likelihood ratio statistic was more nearly χ^2 when the test statistic is modified to be:

$$\chi^2(q)_{adj} = \chi^2(q)[1 - \frac{2N + 5 + 4K}{6T}] \tag{28}$$

which is the form of the test statistic produced by most software packages. Geweke and Singleton(1980) found that the chi-squared approximation (28) is quite good and that the statistic has good power even in modest sample sizes. Most empirical examinations of the **APT** have employed one of the forms of this test statistic.

The problem with the likelihood ratio statistic is that it tests the appropriateness of the *statistical* factor analysis model and not that of the *approximate* factor structure:

$$\Sigma = BB' + \Omega \tag{29}$$

where Ω need not be diagonal due to the presence of residual correlations such as industry effects. The statistical factor analysis model with the diagonal covariance matrix for the idiosyncratic disturbances requires that *any* source of covariation among security returns be classified as a factor while the **APT** counts as factors only those which are *pervasive* and affect many security returns. The likelihood ratio statistic cannot distinguish between correlated idiosyncratic risks which are irrelevant for pricing and common factors which help explain expected returns.[30]

Formally, the difficulty with the likelihood ratio statistic is that the ratio:

$$\frac{|BB' + D|}{|BB' + \Omega|} = \frac{|D||I + B'D^{-1}B|}{|\Omega||I + B'\Omega^{-1}B|} \tag{30}$$

is not one. As a consequence, the noncentrality parameter of the test statistic (28) is not equal to zero which causes the statistic to take on larger values than would be expected from a χ^2 random variable. The following simple numerical example illustrates the potential severity of the problem. Suppose that there are K common factors and that the economy consists of many industries, each consisting of two firms. For simplicity, both firms within an industry are assumed to have identical factor loadings, idiosyncratic variances of unity, and correlation between their idiosyncratic disturbances of ρ. The idiosyncratic disturbances are presumed to be uncorrelated across industries and the value of ρ is taken to be identical for each. In this example, the ratio (30) takes on the value:[31]

$$\frac{|BB' + D|}{|BB' + \Omega|} = \frac{\prod_{j=1}^{K}(1 + \sum_{i=1}^{N} B_{ij}^2)}{(1-\rho^2)^{N/2}\prod_{j=1}^{K}(1 + \frac{1-\rho}{1-\rho^2}\sum_{i=1}^{N} B_{ij}^2)} \approx \frac{1}{(1-\rho^2)^{N/2}(\frac{1-\rho}{1-\rho^2})} \tag{31}$$

If the number of firms in the sample is 750 (the cross-section sizes that we work with) and the correlation between the idiosyncratic disturbances within industries is 0.9 the ratio is virtually infinite (our computer reports that its inverse is zero in double precision), suggesting that we would nearly always reject the null hypothesis that K common factors underlie security returns. Even if the typical industry effect correlation were only 0.1, the ratio would be 43.33 which is large enough to lead to frequent rejection of the null hypothesis.

Of course, this discussion does not suggest that the number of factors cannot be determined by statistical means but rather that this test cannot provide a reliable answer to

[30] This is the essence of the exchange on this subject between Dhrymes, Friend, and Gultekin(1984) and Roll and Ross(1984).

[31] We have implicitly assumed the factor loadings have been normalized so that $B'B$ is a diagonal matrix.

this question. The number of factors can, in principle, be inferred from the tests of the **APT**—these tests are joint tests of the ability of the theory to explain expected returns and the hypothesis that there are K common factors. If the model is rejected with K^* factors but is not rejected based on $K > K^*$ factors,[32] it seems reasonable to conclude that the **APT** is true and that there are K common factors. Unfortunately, such a test may not be powerful in reality but it suggests that the inability of the likelihood ratio test to correctly detect the number of factors in the presence of residual correlations such as industry effects is not fatal. We will find, in fact, that this joint test does not provide much information on the number of factors underlying security returns.

The empirical content of the **APT** lies in the restriction that all securities and portfolios expected returns are spanned by their factor loadings. This suggests a simple test of the appropriateness of a K factor model of security returns as opposed to a K^* factor model where $K^* > K$. Consider regressing the returns on the K^* basis portfolios $\underline{\tilde{R}}^*_{mt}$ on the K basis portfolios $\underline{\tilde{R}}_{mt}$ and a constant term. If the **APT** is true and there are K common factors, the intercepts should be insignificantly different from zero. If, on the other hand, the **APT** is true and there are K^* factors, the returns on the K^* basis portfolios $\underline{\tilde{R}}^*_{mt}$ will embody priced risk factors not contained in $\underline{\tilde{R}}_{mt}$ and hence a joint test that all the intercepts are zero should be rejected. Consequently, our final test will be this test on the intercepts in the following multivariate regression:

$$\underline{\tilde{R}}^*_{mt} = \underline{\hat{\alpha}}^* + \hat{B}^* \underline{\tilde{R}}_{mt} + \underline{\tilde{\epsilon}}^*_t \tag{32}$$

where $\underline{\hat{\alpha}}^*$ is the vector of intercepts in this regression, B^* is the matrix of factor loadings and $\underline{\epsilon}^*_t$ is the associated vector of idiosyncratic disturbances which have covariance matrix Ω^*. We will presume that Ω^* is nonsingular, an assumption which may not be reasonable in all circumstances. If $\underline{\tilde{R}}_{mt}$ and $\underline{\tilde{\epsilon}}^*_t$ are jointly normally distributed, the null hypothesis that the estimates $\underline{\hat{\alpha}}^*$ are insignificantly different from zero can be tested by the usual F statistic:

$$\frac{\underline{\hat{\alpha}}^{*\prime} {\Omega^*}^{-1} \underline{\hat{\alpha}}^*}{1 + \underline{\overline{R}}_m' \hat{\Sigma}_m^{-1} \underline{\overline{R}}_m} \left[\frac{T(T-K-K^*)}{K^*(T-K-1)}\right] \qquad \sim F(K^*, T-K-K^*) \tag{33}$$

[32] Assuming the appropriate in-sample and out-of-sample tests are conducted to guard against overfitting.

under the assumption that the returns $\underline{\tilde{R}}_{mt}$ on the K basis portfolios are measured without error.

What are the power characteristics of this test for the number of factors? Unfortunately, they are not good when the basis portfolios contain some, but not much, idiosyncratic risk, the case that is probably the most likely to occur in practice. The problem is most easily seen in the case where the true number of factors is K^* and the basis portfolios $\underline{\tilde{R}}^*_{mt}$ are measured without error. In this case, the *true* relationship between $\underline{\tilde{R}}^*_{mt}$ and $\underline{\tilde{R}}_{mt}$ is given by:

$$\underline{\tilde{R}}_{mt} = B_m \underline{\tilde{R}}^*_{mt} + \underline{\tilde{\epsilon}}_{mt} \tag{34}$$

where B_m is a matrix of factor loadings and $\underline{\tilde{\epsilon}}_{mt}$ is the associated vector of idiosyncratic disturbances which are assumed to have nonsingular covariance matrix Ω_m. If this is the correct model, then the intercepts $\underline{\alpha}^*$ in (32) satisfy:

$$\underline{\alpha}^* = [I - \Sigma^*_m B'_m \Sigma^{-1}_m B_m]\overline{\underline{R}}^*_m \tag{35}$$

As a consequence, the quadratic form which is of central interest will satisfy:

$$\underline{\alpha}^{*'} \Omega^{*^{-1}} \underline{\alpha}^* = \overline{\underline{R}}^{*'}_m [\Sigma^*_m + \Sigma^*_m B'_m \Omega^{-1}_m B_m \Sigma^*_m]^{-1} \overline{\underline{R}}^*_m \tag{36}$$

What happens when the elements of Ω_m are small as is likely to be the case when the $\underline{\tilde{R}}_{mt}$ are the returns on well-diversified portfolios? In this circumstance, the inverse of Ω_m is likely to be large relative to the other parameters. Hence, the quadratic form (36) will tend to be small, rendering unlikely the rejection of the null hypothesis when it is false. In consequence, this test, like the likelihood ratio test, is not likely to provide reliable information regarding the number of factors underlying security returns. If there is a superior procedure, however, for ascertaining the true number of factors, we have not yet found it.

V. Empirical Results

A. Data Considerations

The following subsections detail our results regarding the validity of the **APT**, the comparative merits of the zero beta and riskless rate formulations, and the evidence on the number of factors underlying security returns. However, first we briefly describe our choices concerning the appropriate data for estimation and testing purposes. In particular, we must confront interesting tradeoffs in the choice of the appropriate observation frequency.

The **CRSP** files provide two sets of equity returns: daily returns on all stocks listed on the New York and American Stock Exchanges since July 1962 and monthly returns on all securities listed on the New York Stock Exchange since 1926. The *potential* benefit associated with the use of daily data in the estimation of variances and covariances is enormous since the precision with which these parameters are estimated hinges on the frequency of observation. Of course, enthusiasm for daily data must be tempered by the well-known problems of nontrading and thin trading which bias the estimates of these moments. As shown by Blume and Stambaugh(1983) and Roll(1983), for instance, the seemingly trivial bid-ask spreads in equity returns lead to serious biases in mean returns and, sadly, this bias is *directly* related to the frequency of observation. Moreover, daily data provides no such advantages when estimating mean returns whose precision depends on the length of the estimation interval and not on the frequency of observation.

These observations have obvious relevance for estimating factor models for security returns and for testing the **APT**. Greater precision in the estimates of variances and covariances confers corresponding improvements in the precision of the estimated factor loadings and idiosyncratic variances, the basic inputs into the subsequent analysis. Biases in mean returns can lead to incorrect inferences regarding the validity of the theory. We opted for a compromise solution in the choice of an observation frequency. Following Roll and Ross(1980) and most subsequent empirical investigators, we estimated our factor models for security returns with daily data since we surmised that the gain in precision offset the thin trading biases in the estimation of covariance matrices.[33] The estimated loadings and idiosyncratic variances were then used to form the portfolio weights of the requisite basis portfolios as described in Section III.

We test the theory and its various aspects using weekly returns data which we formed by continuously compounding daily returns from Wednesday to Tuesday. Consequently, ba-

[33] As a check, we also present results based on factor models estimated with weekly and monthly returns. Currently, in Lehmann and Modest(1985c), we are performing a more thorough examination of the appropriate periodicity for estimating factor models and the issues associated with thin trading and temporal aggregation bias.

sis portfolio returns were computed by multiplying the portfolio weights by the corresponding weekly returns on individual securities. Similarly, our testing strategy required returns on the usual market proxies and so we computed weekly returns on the **CRSP** equally-weighted and value-weighted indices by continuously compounding their daily returns in the same fashion. Note that this means that the market proxies that we use contains both **NYSE** and **AMEX** securities unlike the versions which appear on the monthly returns file, which contain only **NYSE** stocks. All of the relevant test statistics were constructed utilizing these weekly returns with one exception.[34] The tests comparing the returns on the orthogonal equity portfolios and the riskfree rate were performed in monthly data since we were unable to obtain returns on one week Treasury bills on a Wednesday to Tuesday basis.

Two other important choices involve the length of the estimation interval and which firms to include in our sample. As noted above, increasing the estimation interval leads to greater precision of the estimated mean returns. However, longer estimation intervals render more unreasonable the assumption of constant factor loadings that is typically required for testing. As a consequence, we assumed stationarity over five year subperiods and divided the time interval covered by the **CRSP** daily returns file into four periods: 1963 – 1967, 1968 – 1972, 1973 – 1977, and 1978 – 1982. Within each period, we excluded securities which were not continuously listed or which had missing returns and ignored the possible selection bias inherent in this strategy. The remaining securities numbered 1001 in the first period, 1359 in the second period, 1346 in the third period, and 1281 in the final five year period. The number of daily observations in these samples totalled 1259, 1234, 1263, and 1264, respectively, while there were 260 weekly observations in each five year period. The **CRSP** daily file (with few exceptions) lists securities in alphabetical order by their most recent name. We randomly reordered the securities in each subperiod to guard against any biases induced by the natural progression of letters (IBM, International Paper, etc.). The usual sample covariance or correlation matrix of these security returns provided the basic input to our subsequent analysis. Each period we estimated five, ten, and fifteen factor models using the first 750 securities in our randomly reordered data file.

[34] We repeated many of our tests in monthly data and verified that the conclusions reported here are robust with respect to this choice.

B. Tests of the APT

Does the **APT** provide a comprehensive explanation of the expected returns of securities listed on the **NYSE** and the **AMEX**? This important question remains unsettled despite the empirical work of many investigators, including Gehr(1976), Roll and Ross(1980), Reinganum(1981b), Hughes(1982), Brown and Weinstein(1983), Chen(1983), Gibbons(1983), Dhrymes, Friend, and Gultekin(1984) and Dhrymes, Friend, Gultekin and Gultekin(1985). A reasonable characterization of the evidence in these papers is that they are, in general, supportive of the **APT** although they are far from conclusive or uniform in this regard. Some of the reasons for the remaining ambiguities were detailed in Section *IV* along with our solutions to some of the problems with previous tests. As a consequence, we surmise that our more powerful tests will resolve some of the outstanding disputes concerning the validity of the theory.

As noted in the previous section, our strategy for testing the **APT** involves examination of the ability of the theory to account for well-documented empirical anomalies which provide the basis for the rejection of the mean-variance efficiency of the usual market proxies. Tables 1 through 6 provide tests based on three such anomalies: (*i*) firm size, (*ii*) dividend yield, and (*iii*) own variance. Table 1 reports on tests using portfolios formed on the basis of market capitalization. The portfolios were formed by ranking the stocks in our sample by the magnitude of their equity market values at the end of the period preceding the test period, splitting the ranked securities into either five or twenty groups consisting of (approximately) equal numbers, and then constructing equally weighted portfolios from the stocks in each group.[35] Tables 2 and 3 provide the same information as Table 1, but serve as checks that the results in Table 1 do not hinge on peculiarities involving thin trading or January. The sole difference between Tables 1 and 2 is that Table 2 presents results when the factor models were estimated using weekly and monthly data as inputs rather than daily data. In a similar spirit, the tests in Table 3 are based on returns that exclude those occurring in January to ensure that we are not convolving the turn of the year and size effects. Table 4 is similar to the first three except that portfolios were formed on the basis of dividend yield in the year preceding the test period. The ranking procedure was

[35] We also performed tests based on ten such portfolios but the results were similar to those obtained with either the five or twenty portfolios and so we omitted them in the interest of space conservation.

somewhat different as well since we formed an equally weighted portfolio of the firms that paid zero dividends and then formed the remaining four or nineteen portfolios by ranking the remaining dividend-paying stocks in the same fashion as described above.[36] Table 5 reports on the tests based on the variances of the returns in the year preceding the test period (computed from daily data) with portfolios formed as outlined above. Recall that all tests were based on weekly returns with our *weeks* running from Wednesday to Tuesday in order to insure that they did not begin and end on nontrading days too often and to mitigate biases caused by the day-of-the week effect. This means that all tests are based on 260 observations except for the size related tests which exclude January returns, which involve 235 observations.

Table 6 merits separate comment. Their appears to be widespread concern that the **APT** cannot be rejected given the freedom to extract an arbitrary number of factors. While the statement is trivially true if we extract almost as many factors as the number of securities in the analysis, it is false when the number of factors is small since the **APT** predicts that only covariance risk measures explain expected returns not other security characteristics such as firm size, dividend yield, and own variance. Nevertheless, the reasonable question implicit in the concern involves the potential problem of overfitting returns by testing the **APT** with the same securities used to estimate the factor model for security returns. In order to guard against this possibility, we conducted tests similar to those reported above using only securities which were not used to estimate the factor models. This means we formed portfolios based on firm size, dividend yield, and own variance from 251 securities in the first period, 609 securities in the second period, 596 securities in the third period, and 531 securities in the final period. Table 6 only reports results for five portfolios formed from these securities and omits the results for the size tests which exclude January returns as well in order to conserve space.

Each table reports the F statistics for both the riskless rate and zero beta formulations of the **APT** and for five, ten, and fifteen factor models. In addition, they present the large sample F statistics for the mean-variance efficiency of the **CRSP** equally-weighted and value-weighted indices of **NYSE** and **AMEX** stocks.[37] The first half of each table

[36] We also carried out tests by ranking on dividend yield without special treatment of the zero dividend group. The results were very similar to those reported here.

[37] Note that these statistics are formed from the *sorted* portfolios and are subject to

reports the results based on five sorted portfolios and the second half presents those based on twenty such portfolios. The first four rows of each half of the table provide the relevant F statistic for each of the four five year periods: 1963 – 1967, 1968 – 1972, 1973 – 1977, and 1978 – 1982. The subsequent row of each table provides an approximate χ^2 statistic testing the joint significance of the four F statistics. The aggregated statistics are obtained by multiplying the individual F statistics by their numerator degrees of freedom and summing these quantities over the four periods. The resulting test statistic is approximately distributed χ^2 with degrees of freedom equal to the sum of the numerator degrees of freedom of the individual period F statistics. The number reported under each test statistic is the marginal significance level of the test statistic under the null hypothesis (i.e. the probability of obtaining a test statistic at least as large as that obtained when the null hypothesis is true).

The five size-related tests provide sharp evidence against the **APT**. The aggregate χ^2 statistic for the joint significance of their intercepts across sample periods have marginal significance levels below 10^{-3} across both **APT** formulations and the number of factors. Examination of Tables 2 and 3 confirms that this phenomenon does not arise solely from thin trading or January returns since the marginal significance levels are less than 5% in all cases. Perusal of the subperiod results reveals considerable uniformity in the results—virtually all of the corresponding subperiod F statistics which include January returns or correct for thin trading are large enough to reject the **APT** at conventional significance levels irrespective of the number of presumed factors or of the version of the theory and many such statistics computed excluding January returns are large enough to reject in the subperiods as well.

It is also worth noting that these results are not just reflections of unusually large intercepts for the smallest firm (i.e. fifth quintile) portfolio. To be sure, this portfolio has large and positive alphas in all four periods. However, the size effect is not limited to the smallest firms: the fourth quintile appears to plot above the security market hyperplane and the largest firm portfolio consistently plots below the security market hyperplane. This large firm effect receives striking confirmation in regressions of the value-weighted index

the power difficulties discussed in the text. This does not seem to be a problem in this application.

on the basis portfolios which yield highly significant negative intercepts on the order of 2.5% and 4.5% per year for all but the first five year period. While some of this effect may be attributable to nonstationarity associated with the changing weights of the value-weighted index, the effect remains (albeit with lesser magnitude) when we use a fixed weight large firm portfolio. The nature of the small firm effect is further illuminated by similar regressions—the equally-weighted index composed of **NYSE** and **AMEX** stocks has large, positive, and significant alphas in all periods and, in monthly data, the equally-weighted index comprising only **NYSE** securities has economically and statistically insignificant intercepts in all but the second five year period. Not surprisingly, the small firm effect is concentrated in the small firms trading on the **AMEX**.

The five portfolio tests of the mean-variance efficiency of the equally-weighted and value-weighted indices provide similarly striking documentation of the magnitude of the size effect. The aggregate χ^2 statistic rejects the mean-variance efficiency of the equally-weighted index at marginal significance levels below 10^{-9} while the same statistic constructed excluding January returns rejects at marginal significance levels below 10^{-4}. The analogous aggregate marginal significance levels for the tests of the mean-variance efficiency of the value-weighted index are below 10^{-3} for the whole sample and at the 6% level when January returns are excluded. Once again, there is considerable uniformity in the subperiod results since the mean-variance efficiency of both indices is rejected in all but the second period in the whole sample and in the first and fourth periods when January returns are excluded.

The size-related results based on twenty portfolios tell a somewhat different story. The mean-variance efficiency of the usual market proxies is rejected in aggregate at marginal significance levels below 10^{-4} for the equally-weighted index and at the 5% level for the value-weighted index in the whole sample while only the equally-weighted index is rejected (at the 2% level) when January returns are excluded. In contradistinction, only the zero beta version of the five factor model is rejected at the 5% level in the whole sample and no version of the **APT** is rejected when January returns are excluded. Examination of the individual period results suggests that the failure to reject the **APT** is no accident since in all but the second five year period no F statistic attains a marginal significance level below 20%.

Examination of the individual portfolio intercepts does reveal interesting patterns. As would be expected from the regressions of the **CRSP** indices on the basis portfolios suggests, the size effect is concentrated in the extreme size portfolios in all five year periods. Of course, the smallest firm portfolio has much larger positive intercepts in all periods but the small firm effect extends to between three and five other small firm portfolios. Similarly, the large firm effect typically covers the three largest firm portfolios. The **APT** is not rejected with twenty size portfolios because the size effect is concentrated in the largest and smallest firms. This contrasts with the intercepts associated with the **CRSP** indices for which the size effect affects many of the twenty portfolios. This explains why the mean-variance efficiency of both indices is rejected in the twenty portfolio case.[38]

It is also worth noting that our concern over the adverse power consequences associated with basing our tests on different numbers of characteristics-based portfolios seems warranted. Our failure to reject the **APT** based on twenty size portfolios after we rejected the theory based on five size portfolios is suggestive in this regard. Our examination of the individual intercepts confirms the appropriateness of the rejections—there appear to be both small and large firm effects that are not accounted for by our basis portfolios. Moreover, we also experimented with similar tests to those reported for the **CRSP** indices where we estimated the **APT** risk premia with cross-sectional regressions of the size portfolios on the portfolio factor loadings and then employed a large sample F statistic to test the joint significance of the time series intercepts of the size portfolios. In no case were we able to reject the **APT** with this procedure with either ten or twenty size portfolios. Considerable caution is clearly warranted in implementing mean-variance efficiency tests.

Our final observation on size-related tests involves the results reported in Table 6 which exclude the 750 securities which were used to construct the basis portfolios. The mean-variance efficiency of both indices is rejected in this smaller sample of securities with aggregate marginal significance levels less than 10^{-5} for the equally-weighted index and less than 10^{-2} for the value-weighted index. By contrast, only two **APT** formulations are

[38] As noted in Section IV, our results differ markedly from those obtained in Chen(1983). It is possible that this is a consequence of the elastic programming algorithm Chen employed to produce portfolios of small and large firms which had identical sample factor loadings. If the algorithm produced portfolios which placed relatively small weight on the very smallest and largest firms in his sample, the discussion in the text suggests that these portfolios would not exhibit a very large size effect.

rejected at conventional significance levels: the five factor zero beta model was rejected at an aggregate marginal significance level near 2% while the five factor riskless rate version was marginally rejected at just over the 9% level. The rejections of the **CRSP** indices verify that the good performance of the basis portfolios is not due to the absence of power to reject reasonable hypotheses in this sample. The results are probably more a consequence of the *size* of this sample—there are fewer very large and very small market capitalization firms in these subsamples with which to reject the theory.

The tests for the two **CRSP** indices based on dividend-sorted portfolios reject their mean-variance efficiency as would be expected from the earlier work of Litzenberger and Ramaswamy(1979), Blume(1979), and Elton, Gruber, and Rentzler(1983). The aggregate χ^2 statistics based on five portfolios record rejections of the mean-variance efficiency of the equally-weighted index at a marginal significance level below 10^{-10} and of the value-weighted index at just below the 2% level. The corresponding individual period results confirm the aptness of these rejections since the large sample F statistics reject the null hypothesis for both indices at conventional significance levels for all but the second five year period. The mean-variance efficiency of the equally-weighted index is also sharply rejected in the twenty portfolio tests and in the subsample tests reported in Table 6. The tests fail to reject the mean-variance efficiency of the value-weighted index in the twenty portfolio case and marginally reject in the subsample results reported in Table 6. Our examination of the individual portfolio intercepts reveals a well-known pattern: significant positive alphas for the zero dividend and high dividend groups and negative intercepts for the remaining portfolios.

The dividend-related tests lead to very different conclusions regarding the validity of the **APT**. There is very little evidence in Table 4 against the theory. Only the ten and fifteen factor zero beta models are nearly rejected at the 5.5% and 7.6% marginal significance levels, respectively, using five dividend-related portfolios. The only evidence against the theory in the individual subperiods occurs in the first subperiod: both versions of the ten and fifteen factor models are rejected at marginal significance levels between 2% and 4% with five dividend-sorted portfolios. There is no evidence against the **APT** in the remaining five portfolio results and no evidence at all in the twenty portfolio test statistics or those in Table 6 which exclude the 750 securities used to create the basis portfolios.

Moreover, examination of the portfolio intercepts provides little suggestion that the test statistics are missing an important dividend effect.[39]

The results for portfolios based on own variance mirror those obtained in the dividend yield case. The mean-variance efficiency of the **CRSP** indices is rejected at almost identical marginal significance levels in most cases and more sharply for some of the remaining statistics, particularly those relating to the value-weighted index. The **APT** basis portfolios do not yield intercepts that are significant over the whole sample with the exception of the five factor zero beta model which receives a marginal rejection at the 9% level. In addition, many of the F statistics are marginally significant (between the 6% and 10% level) in the first five year period although the remaining test statistics are typically grossly insignificant (many at the 90% level and larger). Once again, the information in the test statistics is a reliable guide to the behavior of the individual portfolio intercepts. The basic message is similar: the rejections of the mean-variance efficiency of both **CRSP** indices suggests that the own variance portfolios have power against reasonable alternatives and the failure to reject the **APT** suggests that the theory provides an adequate account of their risk and return.

C. Comparison of the Riskless Rate and Zero Beta Models

In this section we compare the riskless rate and zero beta formulations of the **APT**. As noted in Section IV, our tests examine two dimensions along which these models differ: (i) the riskless rate interpretation predicts that the orthogonal portfolios constructed from the factor models should earn the riskfree rate while the zero beta model implies that these portfolios should have zero returns and (ii) the sum of the loadings of both individual securities and portfolios should be one if the zero beta formulation is appropriate. The first implication also provides a test of the **APT** itself since these orthogonal portfolios could earn significant negative returns or returns significantly greater than the riskfree rate in violation of the theory. It is certainly true that estimated zero beta rates in a **CAPM** setting are usually significantly greater than the riskfree rate.

[39] In particular, there is some evidence of positive (but usually insignificant) intercepts for the zero dividend and high dividend portfolios. In contrast to the **CAPM** results, the intercepts for the remaining portfolios sometimes have mixed signs and are typically economically and statistically insignificant. We plan to investigate this further in subsequent research.

Table 7 reports summary statistics regarding the sample behavior of the orthogonal portfolios constructed from five, ten, and fifteen factor models as well as those pertaining to fluctuations in the one month Treasury bill rate. Recall that the data underlying this Table are monthly in contrast to the remaining results in this paper. For each orthogonal portfolio, we report the mean return, the standard deviation of its return, and the t-statistic for the hypothesis that the mean return is significantly different from zero as well as the difference in mean return between this portfolio and the one month Treasury bill rate and the t-statistic for this mean return difference. The marginal significance levels of the two t-statistics are reported as well. These statistics are presented for each of the four five year periods in the first four rows. The final row presents two approximate χ^2 statistics along with their marginal significance levels.[40] The first such statistic is for the joint hypothesis that the mean returns on the orthogonal portfolio were jointly significantly different from zero across the four periods while the second χ^2 statistic provides the analogous test for the difference in mean returns between the orthogonal portfolios and one month Treasury bills. Finally, we report summary statistics describing the behavior of the riskless rate. For each of the four subperiods, Table 7 gives the mean, standard deviation, and t statistic (along with its marginal significance level) of returns on one month Treasury bills while the final row provides the approximate χ^2 statistic for the hypothesis that the mean returns are jointly significantly different from zero.

The results in Table 7 suggest considerable uniformity in the behavior of the orthogonal portfolios from the five, ten, and fifteen factor models. The t-statistics for the mean returns on the orthogonal portfolios of all three factor models are highly significant in the final three subperiods although they are insignificant in the first period. The χ^2 statistics for the joint significance of the mean returns for each orthogonal portfolio across the four sample periods have marginal significance levels below 10^{-5} for each factor model. In contradistinction, the corresponding t-statistics for the differences in mean returns between these portfolios and the one month Treasury bill are insignificant in each subperiod with two exceptions: in the second subperiod, the mean return difference for the orthogonal portfolio from the ten factor model is marginally significant at the 7.5% level while that from the fifteen

[40] It proved to be convenient with our software to produce these approximate χ^2 statistics instead of the usual F statistics. Fortunately, with these sample sizes the difference in marginal significance levels would only show up in inconsequential decimal places.

factor model is significant at the 2.7% level. In addition, these mean return differences have mixed signs, negative in the first and fourth periods and positive in the middle two periods. Moreover, the aggregate χ^2 statistics for the joint significance of the mean return differences have marginal significance levels of 0.19 for the five factor model, 0.17 for the ten factor model, and 0.092 for the fifteen factor model. Only the fifteen factor results reflect a marginal rejection of the riskless rate model, a rejection attributable solely to the results from the second subperiod.

The statistics reported in Table 7 provide support for the riskless rate interpretation of the **APT**. The zero beta model receives no such support—the mean returns on the orthogonal portfolios are jointly and, in three of four periods, individually significantly different from zero at low marginal significance levels. The riskless rate interpretation seems to be quite consistent with the data since the only evidence against the model is from the second period for the orthogonal portfolios from the ten and fifteen factor models. Of course, these results may only reflect power problems with these t-statistics although this interpretation appears to be difficult to sustain due to the apparent precision of the estimated means for the final three periods. Similarly, one might be tempted to reject the **APT** since these orthogonal portfolios are not *riskless*, their sample standard deviations are typically ten to twenty times those of returns on one month Treasury bills. This could occur because the **APT** is false or because 750 securities is insufficient to eliminate both factor risk and idiosyncratic risk as discussed in Section IV. In addition, note that our tests are considerably more powerful and more consistent with the riskless rate version of the **APT** than the mixed results obtained by previous authors. Moreover, these results stand in sharp contrast to those obtained in studies of the zero beta **CAPM**, where the estimated zero beta rates are typically significantly greater than the riskless rate.

Of course, there is a second test for the validity of the zero beta **APT**—tests of the hypothesis that the loadings of individual securities or portfolios sum to one. This test could lead to different conclusions than those that follow from Table 7 so that it can shed light on the validity of the **APT** as well. Table 8 reports the relevant test statistics for the hypothesis that the portfolios formed on the basis of firm size, dividend yield, and own variance have loadings that sum to unity. Attention is restricted to the results based on five such portfolios in order to conserve space which involves little sacrifice since the results for ten and twenty portfolios were much less favorable to the zero beta formulation. As noted in Section 4, we present F statistics based on both raw return and excess return regressions. The first four columns present the relevant F statistics for the four subperiods and their marginal significance levels. The final column presents the approximate χ^2 statistic, constructed along the lines of the analogous statistic in Section V.B above, for the joint significance of the four subperiod F statistics along with its marginal significance level.

The results in Table 8 suggest overwhelming rejection of the zero beta formulation of the **APT**. The aggregate test statistics based on excess return regressions reject the null hypothesis at marginal significance levels below 10^{-27} and the least significant subperiod F statistic has a marginal significance level of 2.3%. The hypothesis fares only slightly better in the raw return regressions—the aggregate test statistics for the loadings of the five size portfolios have marginal significance levels of 0.11 and 0.13 for the ten and fifteen factor models, respectively, while that associated with own variance and fifteen factors is significant at the 11% level as well. Of course, the remaining marginal significance levels of the aggregate test statistics indicate overwhelming rejection at conventional significance levels as do many of the subperiod results. Moreover, these results are the *most* favorable to the zero beta model—we would have made no such caveats had we chosen to report the twenty portfolio results. It is certainly hard to imagine a more complete rejection of the zero beta version of the **APT**.

D. Assessing the Number of Factors

The final question we consider involves the number of factors underlying security returns. Our evidence on this question is presented in Tables 9 and 10. As expected, the results are inconclusive. Panel A of Table 9 contains summary χ^2 statistics for the joint significance of the mean returns of the basis portfolios constructed from different factor models. These χ^2 statistics are of potential interest for at least two reasons: one having to do with the validity of the **APT** and the other with the number of factors. First, it is an implication of the **APT** that at least one of the factor risk premia should be different from zero. As is readily apparent, our large cross-sections yielded basis portfolios which had highly significant mean returns in aggregate and in most of the individual subperiods as well. This is in sharp contrast to the frequently insignificant mean returns of basis portfolios constructed from smaller cross-sections in other studies. The second reason these statistics may be of interest is that the marginal significance levels of the χ^2 statistics have some potential for shedding light on the number of factors that embody priced risk.

Why should this be true? Suppose that we estimate two factor models for security returns with K and K^* factors with associated returns $\underline{\tilde{R}}_{mt}$ and $\underline{\tilde{R}}^*_{mt}$. Assume K^* is the true number of factors. The key insight is that the noncentrality parameter associated the usual χ^2 statistic for testing that the mean returns on the K basis portfolios are jointly zero is less than the noncentrality parameter associated with testing that the mean returns on the K^* basis portfolios are significantly different from zero.[41] Since the noncentrality parameter of the true basis portfolios is larger, the χ^2 test statistic of the K^* portfolios should typically be farther out in the tails of its distribution than that of the K portfolios. In consequence, the marginal probabilities of the test statistic for the hypothesis that the mean returns on $\underline{\tilde{R}}^*_{mt}$ are zero ought to be smaller than the corresponding numbers for the mean returns on $\underline{\tilde{R}}_{mt}$. Of course, more precise statements are not possible in this setting in the absence of a more detailed characterization of the joint distribution of the two sets of basis portfolios. Moreover, the ranking of the marginal probabilities need not obtain in any period due to sampling error. Nevertheless, aggregation of the relevant test statistics might mitigate some of the harmful effects of sampling error and might yield insights into the number of factors underlying security returns.

Unfortunately, the χ^2 statistics for the hypothesis that basis portfolio mean returns are jointly significantly different from zero are inconclusive. While the aggregate χ^2 statistics suggest that the ten factor model yields the basis portfolios with the most significant mean returns, examination of the subperiod results shows that the basis portfolios constructed from the five factor model had the most significant mean returns in the first two subperiods and those associated with the ten factor model had the lowest marginal significance levels in the latter two five year periods. This may be interpreted as very weak evidence in favor of a ten factor model.

The remaining results are equally uninformative. Panel B of Table 9 displays the

[41] In terms of the notation of Section IV.C, $\overline{\underline{R}}'_m \Sigma_m^{-1} \overline{\underline{R}}_m$ is the noncentrality parameter of the K basis portfolios which can be rewritten as $\overline{\underline{R}}^{*\prime}_m B'_m [B_m \Sigma^*_m B'_m + \Omega_m]^{-1} B_m \overline{\underline{R}}^*_m$ $= \overline{\underline{R}}^{*\prime}_m [\Sigma^*_m + (B'_m \Omega_m^{-1} B_m)^{-1}]^{-1} \overline{\underline{R}}^*_m$. The noncentrality parameter of the K^* basis portfolios is $\overline{\underline{R}}^{*\prime}_m \Sigma^{*\,-1}_m \overline{\underline{R}}^*_m$.

usual likelihood ratio statistics for the number of factors. As anticipated, the likelihood ratio statistics for the number of factors overwhelmingly reject the hypothesis that five, ten, or fifteen factors are sufficient to explain covariation among security returns when the idiosyncratic disturbances are presumed to be uncorrelated across securities. Finally, our tests in Table 10 for the joint significance of the intercepts from the regression of one set of basis portfolios on another provides no evidence against the riskless rate version of the **APT** coupled with any number of factors from five to fifteen. Only the joint hypothesis that the five factor zero beta model is correct is rejected at the 3.7% level by the basis portfolios from the ten factor model.

Tables 9 and 10 are, in some respects, the least satisfying in the paper. They provide very little information regarding the number of factors which underlie the **APT**. As the analysis in Section IV.D suggests, our tests have little power to discriminate among models with different numbers of factors. The likelihood ratio cannot tell the difference between pervasive common factors and nonpriced industry effects and so the sharp rejections may reflect the inadequacy of these models or the inappropriateness of the assumption that the idiosyncratic disturbances of different firms are uncorrelated. The marginal differences in the marginal significance levels of basis portfolio mean returns seem to be equally uninformative. The regression tests provide no evidence against any of the factor models but the analysis in Section IV suggests that the tests have little power.

As a consequence, there are two plausible readings of the evidence we have examined. One is that there is no real evidence against a five factor model providing an adequate empirical basis for the **APT**. The other interpretation is that there is no adequate basis for choosing among factor models and we instead must rely on intuitions regarding the comparative performance of different factor models. Our hunch is that, if a five factor model is not appropriate, a ten factor model is sufficient and that there is no need to move to fifteen factors.

VI. Conclusion

This paper has been devoted to the accumulation of facts and the sifting of evidence regarding the validity of the **APT** in its various incarnations. In this pursuit, we have reached several firm conclusions and have left some issues largely unresolved. In particular, our empirical implementation of the theory proved incapable of explaining expected returns on portfolios composed of securities with different market capitalizations although it provided an adequate account of the expected returns of portfolios formed on the basis of dividend yield and own variance. In addition, it appears that the zero beta version of the **APT** is sharply rejected in favor of the riskless rate model and that there is little basis for discriminating among five and ten factor versions of the theory.

The sharpest evidence we obtained concerns the comparative merits of the zero beta and riskless rate versions of the **APT**. The implications of the two models differ in two dimensions: (i) the zero beta model requires the factor loadings of both securities and portfolios sum to one and (ii) the riskless rate formulation predicts that the intercept in the **APT** pricing relation is the riskless rate while the zero beta version implies that it is zero. Our tests of the first implication sharply reject the hypothesis that the loadings of portfolios based on firm size, dividend yield, and own variance sum to unity at arbitrarily low marginal significance levels. Moreover, our examination of the intercept λ_0 in the pricing relation confirmed the appropriateness of these rejections as they proved to be significantly different from zero at low marginal significance levels in three of four periods and in aggregate. The **APT** and, in particular, the riskless rate version of the model received additional support in these tests in that these intercepts proved to be insignificantly different from the riskfree rate in aggregate and in all but one subperiod. This is somewhat surprising given that zero beta rates in **CAPM** studies are typically significantly greater than riskfree rates.[42]

Considerable ambiguity remains regarding the number of common factors underlying security returns. This is not surprising in that the analysis in Section IV failed to turn up a test which could reliably discriminate among alternative factor models. The evidence presented in Section V is consistent with either the five, ten or fifteen factor model. In light of the similar performance of the ten and fifteen factor models in most instances, we conjecture that five or ten factors is sufficient if the **APT** is true.

By far the most interesting results in the paper concern the validity of the **APT** itself. The **APT** fared well when confronted with the strong relationship between average returns and dividend yield and own variance. The **APT** provides an adequate account of their risk and return where risk adjustment with the **CAPM** with the usual market

[42] Of course, this does not constitute evidence against the zero beta **CAPM**.

proxies fails. This is noteworthy since the **APT** provides a risk based explanation of these phenomena in contrast to the usual tax related explanation of the dividend effect and the transactions cost account of the relationship between own variance and average returns.[43] In contradistinction, the tests based on market capitalization provide sharp evidence against the **APT**, although the form of the size effect appears different from that documented in **CAPM** studies.

How should we interpret this failure to account for the size effect? One possibility is that despite our cautious attentiveness to the statistical underpinnings of this analysis, our procedures proved incapable of overcoming measurement error caused by our inadequate sample size, asynchronous trading, or any of the other problems discussed in Section IV. We are persuaded, though, that the large cross-sections that we employ largely mitigate the impact of measurement error. Similarly, the thin trading corrections made in Section V yield no suggestion that the size-related results are attributable to this problem. Moreover, the sharpness of the rejections reported in Tables 1 – 3 suggests that they cannot be attributed to peculiar small sample properties of the test statistics such as those that might result, for example, from non-normality. These considerations suggest that the failure of the **APT** to account for the size effect is credible.

The most obvious interpretation is that we have sharply rejected the **APT**. The ability of a measure of unsystematic risk to successfully explain risk-adjusted returns violates the theory. The analysis above suggests that the rejections are both sharp and believable. This represents a clear failure of our empirical implementation of the **APT**.

The concentration of the size effect in the very smallest and largest firms, however, suggests an alternative explanation of these results. Suppose that there is a *small firm* factor in that the business cycle risk of small capitalization firms is much greater than that of better capitalized firms. In addition, suppose that the exposure to this source of risk is small for listed equities but that the risk premium for this factor is large. In particular, suppose that the firms which suffer from significant exposure to this source of systematic risk are *primarily* traded over the counter or are closely held. In these circumstances, our

[43] It is possible that the absence of a measured dividend effect in our **APT** results is consistent with the tax story. This could occur, for example, if one of the risk factors reflected random marginal tax rates impinging on asset pricing and the corresponding factor loadings are the dividend yields of the individual securities.

factor analysis would fail to measure this factor well since few firms in our cross-section would be materially affected by it. Similarly, our rejections of the **APT** are consonant with the smallest market capitalization firms having small positive loadings and the large firms having small negative loadings coupled with a large risk premium for the factor. [44] Hence, this account of the size effect involves measurement error in the factors, measurement error that follows from the assets selected for the analysis rather than as a consequence of our statistical procedures.

The size and the turn-of-the-year effect have thus far evaded a satisfactory risk based explanation. It is worth emphasizing that *our* size effect is largely concentrated in the firms with the largest and smallest market capitalizations which suggests that the **APT** is pricing most listed equities with little error. To paraphrase Henry IV of France to the ambassador Don Pedro of Spain, "Do you mean to say your theory hasn't enough virtues to afford some faults?"

[44] It is suggestive to note that business failure rates rise sharply during recessions and that few of these failures occur among firms listed on the **NYSE** and **AMEX**. This could occur if, for example, credit rationing occurs during recessions and the capitalization of listed equities and their access to credit is sufficient to ride out most recessions. We would expect very large firms to have negative loadings in these circumstances since they could potentially profit from acquisitions obtained during recessions.

TABLE 1: TESTS OF THE APT MEAN RESTRICTION BASED ON SIZE PORTFOLIOS: F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS
(p-values in parentheses)

PERIODICITY OF DATA USED TO ESTIMATE FACTOR LOADINGS: DAILY

NUMBER OF PORTFOLIOS	SAMPLE PERIOD	CAPM EQUALLY WEIGHTED	CAPM VALUE WEIGHTED	FIVE FACTORS ZERO BETA	FIVE FACTORS RISKLESS RATE	TEN FACTORS ZERO BETA	TEN FACTORS RISKLESS RATE	FIFTEEN FACTORS ZERO BETA	FIFTEEN FACTORS RISKLESS RATE
FIVE	January 1963-December 1967	9.17 (.62E-06)	3.18 (.01)	2.33 (.04)	2.45 (.03)	3.04 (.01)	2.88 (.02)	2.82 (.02)	2.74 (.02)
	January 1968-December 1972	1.25 (.29)	.62 (.63)	4.11 (.001)	2.92 (.01)	4.04 (.002)	4.18 (.001)	2.60 (.03)	3.10 (.01)
	January 1973-December 1977	3.28 (.01)	3.12 (.02)	2.27 (.05)	3.00 (.01)	1.36 (.24)	1.84 (.11)	1.56 (.17)	2.04 (.07)
	January 1978-December 1982	6.50 (.54E-04)	2.96 (.02)	3.67 (.003)	2.96 (.01)	2.56 (.03)	2.36 (.04)	2.86 (.02)	2.41 (.04)
	AGGREGATE: January 1963-December 1982	80.77 (.12E-09)	39.60 (.89E-03)	61.88 (.36E-05)	56.68 (.23E-04)	54.99 (.41E-04)	56.31 (.26E-04)	49.16 (.29E-03)	51.46 (.14E-03)
TWENTY	January 1963-December 1967	2.61 (.41E-03)	1.45 (.10)	1.11 (.34)	1.17 (.28)	1.24 (.22)	1.22 (.24)	1.09 (.36)	1.11 (.34)
	January 1968-December 1972	1.46 (.10)	1.36 (.15)	1.99 (.01)	1.44 (.11)	1.81 (.02)	1.56 (.06)	1.39 (.13)	1.22 (.24)
	January 1973-December 1977	1.09 (.36)	1.41 (.12)	1.02 (.43)	1.20 (.26)	.79 (.72)	.96 (.52)	.86 (.64)	.99 (.48)
	January 1978-December 1982	1.77 (.03)	.94 (.53)	1.13 (.32)	.99 (.48)	.81 (.70)	.79 (.73)	.90 (.59)	.77 (.75)
	AGGREGATE: January 1963-December 1982	131.75 (.77E-04)	98.05 (.05)	105.27 (.03)	95.82 (.11)	93.22 (.15)	93.91 (.14)	84.76 (.34)	81.89 (.42)

TABLE 2: TESTS OF THE APT MEAN RESTRICTION BASED ON FIVE SIZE PORTFOLIOS: F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS

(p-values in parentheses)

DATA USED TO ESTIMATE FACTOR LOADINGS	SAMPLE PERIOD	CAPM		FIVE FACTORS		TEN FACTORS		FIFTEEN FACTORS	
		EQUALLY WEIGHTED	VALUE WEIGHTED	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE
WEEKLY	January 1963-December 1967	9.17 (.62E-06)	3.18 (.01)	3.32 (.64E-02)	3.10 (.99E-02)	4.58 (.52E-03)	3.47 (.48E-02)	N/A	N/A
	January 1968-December 1972	1.25 (.29)	.62 (.63)	3.79 (.25E-02)	3.07 (.01)	3.30 (.67E-02)	3.72 (.29E-02)	N/A	N/A
	January 1973-December 1977	3.28 (.01)	3.12 (.02)	4.32 (.86E-03)	3.61 (.36E-02)	3.45 (.49E-02)	4.20 (.11E-02)	N/A	N/A
	January 1978-December 1982	6.50 (.54E-04)	2.96 (.02)	4.82 (.31E-03)	4.21 (.11E-02)	3.52 (.44E-02)	3.38 (.57E-02)	N/A	N/A
	AGGREGATE: January 1963-December 1982	80.77 (.12E-09)	39.60 (.89E-03)	81.25 (.24E-08)	69.95 (.19E-06)	74.25 (.36E-07)	73.85 (.42E-07)	N/A	N/A
MONTHLY	January 1963-December 1967	9.17 (.62E-06)	3.18 (.01)	3.29 (.68E-02)	3.47 (.48E-02)	2.80 (.02)	3.36 (.59E-02)	N/A	N/A
	January 1968-December 1972	1.25 (.29)	.62 (.63)	2.00 (.08)	2.80 (.02)	1.77 (.12)	2.89 (.01)	N/A	N/A
	January 1973-December 1977	3.28 (.01)	3.12 (.02)	3.31 (.66E-02)	2.83 (.02)	2.47 (.03)	2.26 (.05)	N/A	N/A
	January 1978-December 1982	6.50 (.54E-04)	2.96 (.02)	4.04 (.15E-02)	2.32 (.04)	3.48 (.47E-02)	2.99 (.01)	N/A	N/A
	AGGREGATE: January 1963-December 1982	80.77 (.12E-09)	39.60 (.89E-03)	63.20 (.23E-05)	57.10 (.20E-04)	52.60 (.93E-04)	57.50 (.17E-04)	N/A	N/A

TABLE 3: TESTS OF THE APT MEAN RESTRICTION BASED ON SIZE PORTFOLIOS EXCLUDING JANUARY: F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS

(p-values in parentheses)

NUMBER OF PORTFOLIOS	SAMPLE PERIOD	CAPM EQUALLY WEIGHTED	CAPM VALUE WEIGHTED	FIVE FACTORS ZERO BETA	FIVE FACTORS RISKLESS RATE	TEN FACTORS ZERO BETA	TEN FACTORS RISKLESS RATE	FIFTEEN FACTORS ZERO BETA	FIFTEEN FACTORS RISKLESS RATE
FIVE	January 1963- December 1967	6.15 (.10E-03)	1.98 (.10)	1.37 (.24)	1.76 (.12)	2.06 (.07)	1.81 (.11)	2.06 (.07)	1.89 (.10)
	January 1968- December 1972	.26 (.91)	.20 (.94)	2.01 (.08)	1.22 (.30)	1.68 (.14)	1.80 (.11)	1.01 (.41)	1.23 (.29)
	January 1973- December 1977	.87 (.48)	1.56 (.19)	1.47 (.20)	2.12 (.06)	.89 (.49)	1.12 (.35)	.87 (.50)	1.08 (.37)
	January 1978- December 1982	4.44 (.002)	2.38 (.05)	3.56 (.004)	3.45 (.005)	2.22 (.05)	2.29 (.05)	2.52 (.03)	2.21 (.06)
	AGGREGATE: January 1963- December 1982	46.83 (.72E-04)	30.57 (.06)	42.08 (.003)	42.76 (.002)	34.24 (.03)	35.14 (.02)	32.26 (.04)	32.06 (.04)
TWENTY	January 1963- December 1967	1.94 (.01)	1.08 (.37)	.87 (.63)	.94 (.54)	.95 (.53)	.88 (.61)	.85 (.65)	.81 (.71)
	January 1968- December 1972	1.40 (.13)	1.51 (.09)	1.54 (.07)	1.12 (.33)	1.37 (.14)	1.26 (.12)	1.17 (.28)	1.04 (.41)
	January 1973- December 1977	.76 (.75)	1.01 (.45)	.93 (.55)	1.12 (.33)	.78 (.73)	.87 (.63)	.70 (.82)	.74 (.78)
	January 1978- December 1982	1.36 (.15)	.79 (.72)	1.13 (.32)	1.07 (.39)	.77 (.74)	.78 (.74)	.92 (.57)	.78 (.73)
	AGGREGATE: January 1963- December 1982	103.77 (.02)	83.35 (.26)	89.42 (.22)	84.89 (.33)	77.55 (.56)	75.66 (.62)	72.79 (.70)	67.47 (.84)

TABLE 4: TESTS OF THE APT MEAN RESTRICTION BASED ON DIVIDEND YIELD PORTFOLIOS: F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS

(p-values in parentheses)

NUMBER OF PORTFOLIOS	SAMPLE PERIOD	CAPM		FIVE FACTORS		TEN FACTORS		FIFTEEN FACTORS	
		EQUALLY WEIGHTED	VALUE WEIGHTED	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE
FIVE	January 1963- December 1967	9.54 (.33E-06)	2.18 (.07)	1.47 (.20)	1.44 (.21)	2.67 (.02)	2.44 (.04)	2.65 (.02)	2.51 (.03)
	January 1968- December 1972	.96 (.43)	.26 (.90)	1.70 (.14)	.80 (.55)	1.10 (.36)	.58 (.72)	.64 (.67)	.48 (.79)
	January 1973- December 1977	3.43 (.009)	2.64 (.04)	1.56 (.17)	1.76 (.12)	.85 (.52)	.97 (.44)	1.07 (.38)	1.15 (.33)
	January 1978- December 1982	7.58 (.88E-05)	2.46 (.05)	.77 (.57)	1.20 (.31)	1.58 (.17)	1.12 (.35)	1.56 (.17)	1.05 (.39)
	AGGREGATE: January 1963- December 1982	86.08 (.13E-10)	30.15 (.02)	27.48 (.12)	25.99 (.17)	31.01 (.06)	25.54 (.18)	29.62 (.08)	25.96 (.17)
TWENTY	January 1963- December 1967	2.40 (.001)	1.13 (.32)	.91 (.57)	.92 (.56)	1.19 (.26)	1.12 (.33)	1.15 (.30)	1.12 (.33)
	January 1968- December 1972	.55 (.94)	.44 (.98)	1.11 (.35)	.62 (.89)	.96 (.52)	.54 (.95)	.75 (.77)	.52 (.96)
	January 1973- December 1977	1.21 (.25)	1.03 (.43)	.80 (.71)	.73 (.79)	.55 (.94)	.55 (.94)	.54 (.95)	.51 (.96)
	January 1978- December 1982	2.56 (.51E-03)	1.02 (.44)	.69 (.84)	.77 (.75)	.70 (.83)	.67 (.85)	.68 (.85)	.64 (.88)
	AGGREGATE: January 1963- December 1982	127.78 (.19E-03)	68.72 (.71)	70.11 (.78)	61.01 (.94)	67.87 (.83)	57.69 (.97)	62.32 (.93)	55.59 (.98)

TABLE 5: TESTS OF THE APT MEAN RESTRICTION BASED ON OWN VARIANCE PORTFOLIOS: F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS

(p-values in parentheses)

NUMBER OF PORTFOLIOS	SAMPLE PERIOD	CAPM EQUALLY WEIGHTED	CAPM VALUE WEIGHTED	FIVE FACTORS ZERO BETA	FIVE FACTORS RISKLESS RATE	TEN FACTORS ZERO BETA	TEN FACTORS RISKLESS RATE	FIFTEEN FACTORS ZERO BETA	FIFTEEN FACTORS RISKLESS RATE
FIVE	January 1963-December 1967	9.05 (.76E-06)	3.27 (.01)	1.70 (.13)	1.70 (.14)	2.15 (.06)	1.90 (.10)	1.96 (.09)	1.81 (.11)
	January 1968-December 1972	1.14 (.34)	.35 (.84)	1.38 (.23)	.53 (.75)	.74 (.60)	.48 (.79)	.73 (.60)	.67 (.65)
	January 1973-December 1977	2.39 (.05)	1.56 (.19)	1.55 (.18)	1.72 (.13)	.51 (.77)	.89 (.49)	.63 (.68)	.91 (.47)
	January 1978-December 1982	6.86 (.30E-04)	2.36 (.05)	1.15 (.34)	.65 (.66)	.36 (.88)	.39 (.85)	.25 (.94)	.32 (.90)
	AGGREGATE: January 1963-December 1982	81.71 (.82E-10)	30.17 (.02)	28.88 (.09)	22.96 (.29)	18.80 (.53)	18.30 (.57)	17.84 (.60)	18.54 (.55)
TWENTY	January 1963-December 1967	2.75 (.19E-03)	1.82 (.02)	1.55 (.07)	1.55 (.07)	1.50 (.08)	1.39 (.13)	1.35 (.15)	1.28 (.19)
	January 1968-December 1972	.93 (.55)	.78 (.73)	.77 (.75)	.47 (.98)	.60 (.91)	.48 (.97)	.62 (.90)	.54 (.95)
	January 1973-December 1977	1.10 (.35)	.98 (.49)	.75 (.77)	.76 (.76)	.48 (.97)	.52 (.96)	.50 (.97)	.50 (.96)
	January 1978-December 1982	1.95 (.01)	1.03 (.43)	.70 (.82)	.62 (.90)	.49 (.97)	.49 (.97)	.56 (.93)	.48 (.97)
	AGGREGATE: January 1963-December 1982	127.94 (.18E-03)	87.38 (.18)	75.50 (.62)	67.98 (.83)	61.20 (.94)	57.75 (.97)	60.42 (.95)	56.18 (.98)

TABLE 6: TESTS OF THE APT MEAN RESTRICTION BASED ON FIVE SORTED CONTROL PORTFOLIOS: F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS
(p-values in parentheses)

SORTING CHARACTERISTIC	SAMPLE PERIOD	CAPM		FIVE FACTORS		TEN FACTORS		FIFTEEN FACTORS	
		EQUALLY WEIGHTED	VALUE WEIGHTED	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE
SIZE	January 1963-December 1967	4.69 (.001)	2.79 (.03)	1.40 (.23)	1.37 (.24)	1.49 (.19)	1.43 (.22)	1.26 (.28)	1.25 (.29)
	January 1968-December 1972	1.24 (.29)	1.08 (.37)	2.56 (.03)	1.80 (.11)	1.97 (.08)	1.92 (.09)	1.41 (.22)	1.46 (.20)
	January 1973-December 1977	2.80 (.03)	2.57 (.04)	1.54 (.18)	1.83 (.11)	.77 (.57)	1.03 (.40)	.78 (.56)	1.08 (.37)
	January 1978-December 1982	5.06 (.61E-03)	1.85 (.12)	1.41 (.22)	.75 (.58)	1.03 (.40)	.76 (.58)	1.13 (.34)	.81 (.54)
	AGGREGATE: January 1963-December 1982	55.19 (.33E-05)	33.19 (.007)	34.55 (.02)	28.77 (.09)	26.34 (.15)	25.72 (.18)	22.92 (.29)	23.02 (.29)
DIVIDEND YIELD	January 1963-December 1967	4.67 (.001)	1.18 (.32)	1.67 (.14)	1.59 (.16)	1.61 (.16)	1.40 (.22)	1.39 (.23)	1.25 (.28)
	January 1968-December 1972	.24 (.91)	.37 (.83)	.73 (.60)	.29 (.92)	.38 (.86)	.18 (.97)	.24 (.94)	.20 (.96)
	January 1973-December 1977	2.81 (.03)	2.51 (.04)	.63 (.68)	.73 (.61)	.36 (.88)	.32 (.90)	.40 (.85)	.33 (.89)
	January 1978-December 1982	4.75 (.001)	1.98 (.10)	.68 (.64)	.64 (.67)	1.34 (.25)	.92 (.47)	1.28 (.27)	.85 (.52)
	AGGREGATE: January 1963-December 1982	49.86 (.24E-04)	24.11 (.09)	18.54 (.55)	16.20 (.70)	18.47 (.56)	14.10 (.83)	16.56 (.68)	13.17 (.87)

TABLE 6 (cont'd)

SORTING CHARACTERISTIC	SAMPLE PERIOD	CAPM		FIVE FACTORS		TEN FACTORS		FIFTEEN FACTORS	
		EQUALLY WEIGHTED	VALUE WEIGHTED	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE	ZERO BETA	RISKLESS RATE
	January 1963- December 1967	5.47 (.31E-03)	4.05 (.33E-02)	1.32 (.25)	1.35 (.25)	1.49 (.19)	1.49 (.19)	1.23 (.30)	1.23 (.29)
	January 1968- December 1972	.67 (.61)	.72 (.58)	1.23 (.30)	.47 (.80)	.84 (.52)	.39 (.86)	.46 (.80)	.34 (.89)
OWN VARIANCE	January 1973- December 1977	1.87 (.12)	1.46 (.21)	1.28 (.27)	1.31 (.26)	.36 (.88)	.49 (.78)	.45 (.81)	.53 (.76)
	January 1978- December 1982	5.47 (.31E-03)	2.30 (.06)	.79 (.57)	.57 (.72)	.63 (.67)	.42 (.83)	.48 (.79)	.34 (.89)
	AGGREGATE: January 1963- December 1982	53.92 (.53E-05)	34.12 (.52E-02)	23.13 (.28)	18.49 (.55)	16.58 (.68)	13.95 (.83)	13.13 (.87)	12.21 (.91)

TABLE 7: SUMMARY STATISTICS OF MONTHLY RETURNS ON THE ORTHOGONAL PORTFOLIOS

(p-values in parentheses)

SAMPLE PERIOD	SUMMARY STATISTICS	TREASURY BILL RETURNS	FIVE FACTOR MODEL		TEN FACTOR MODEL		FIFTEEN FACTOR MODEL	
			RAW RETURNS	EXCESS RETURNS OVER T-BILLS	RAW RETURNS	EXCESS RETURNS OVER T-BILLS	RAW RETURNS	EXCESS RETURNS OVER T-BILLS
January 1963-December 1967	Mean	.0033	-.00055	-.0039	.0010	-.0023	.0012	-.0021
	Standard Deviation	.00056	.018	-	.014	-	.014	-
	T-Statistic	45.95 (0.00)	.22 (.82)	1.59 (.11)	.59 (.56)	1.31 (.19)	.71 (.48)	1.19 (.23)
January 1968-December 1972	Mean	.0046	.0084	.0039	.0088	.0043	.0091	.0045
	Standard Deviation	.0011	.022	-	.018	-	.016	-
	T-Statistic	33.65 (0.00)	3.00 (.0027)	1.36 (.17)	3.70 (.00021)	1.78 (.075)	4.45 (.84E-05)	2.21 (.027)
January 1973-December 1977	Mean	.0053	.0081	.0028	.0082	.0029	.0081	.0028
	Standard Deviation	.0011	.022	-	.020	-	.020	-
	T-Statistic	36.25 (0.00)	2.91 (.0036)	1.01 (.31)	3.13 (.0017)	1.12 (.26)	3.11 (.0019)	1.08 (.28)
January 1978-December 1982	Mean	.0088	.0064	-.0024	.0075	-.0013	.0068	-.0020
	Standard Deviation	.0023	.022	-	.022	-	.021	-
	T-Statistic	29.36 (0.00)	2.21 (.027)	.82 (.42)	2.69 (.0071)	.47 (.64)	2.47 (.013)	.73 (.47)
January 1963-December 1982	Mean	.0055	.0056	.0001	.0064	.0009	.0063	.0008
	Joint Chi-Squared Statistic	5420.17 (0.00)	22.41 (.00017)	6.08 (.19)	31.13 (.29E-05)	6.36 (.17)	36.12 (.27E-06)	7.99 (.092)

TABLE 8: TESTS OF THE PARAMETER RESTRICTIONS IMPLIED BY THE ZERO BETA VERSION OF THE APT BASED ON FIVE SORTED PORTFOLIOS: F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS
(p-values in parentheses)

SORTING CHARACTERISTICS	SAMPLE PERIOD	FIVE FACTORS		TEN FACTORS		FIFTEEN FACTORS	
		RAW RETURNS	EXCESS RETURNS	RAW RETURNS	EXCESS RETURNS	RAW RETURNS	EXCESS RETURNS
SIZE	January 1963-December 1967	9.66 (.19E-07)	25.08 (0.00)	1.12 (.35)	4.94 (.25E-03)	1.00 (.42)	4.91 (.27E-03)
	January 1968-December 1972	6.02 (.28E-04)	32.27 (0.00)	2.61 (.025)	2.91 (.014)	1.92 (.092)	4.66 (.44E-03)
	January 1973-December 1977	2.39 (.039)	20.07 (0.00)	0.93 (.46)	33.27 (0.00)	1.11 (.35)	26.30 (0.00)
	January 1978-December 1982	7.41 (.17E-05)	59.78 (0.00)	0.95 (.45)	14.17 (.35E-11)	1.45 (.21)	2.57 (.027)
	AGGREGATE: January 1963-December 1982	685.98 (0.00)	276.43 (0.00)	192.22 (.39E-29)	127.37 (.12E-16)	28.02 (.11)	27.37 (.13)
DIVIDEND YIELD	January 1963-December 1967	8.94 (.79E-07)	14.14 (.34E-11)	3.70 (.30E-02)	3.42 (.53E-02)	2.54 (.029)	3.47 (.48E-02)
	January 1968-December 1972	4.25 (.10E-02)	28.18 (0.00)	2.13 (.062)	8.00 (.52E-06)	1.18 (.32)	10.29 (.60E-08)
	January 1973-December 1977	2.76 (.019)	4.68 (.42E-03)	2.78 (.018)	29.78 (0.00)	1.35 (.24)	23.41 (0.00)
	January 1978-December 1982	5.93 (.34E-04)	142.04 (0.00)	3.20 (.81E-02)	10.10 (.83E-08)	2.68 (.22E-01)	2.65 (.023)
	AGGREGATE: January 1963-December 1982	109.40 (.25E-13)	945.24 (0.00)	59.07 (.99E-05)	256.54 (0.00)	38.72 (.72E-02)	199.12 (.17E-30)

TABLE 8 (cont'd)

SORTING CHARACTERISTICS	SAMPLE PERIOD	FIVE FACTORS		TEN FACTORS		FIFTEEN FACTORS	
		RAW RETURNS	EXCESS RETURNS	RAW RETURNS	EXCESS RETURNS	RAW RETURNS	EXCESS RETURNS
	January 1963- December 1967	6.01 (.29E-04)	13.70 (.80E-11)	1.41 (.22)	4.96 (.24E-03)	1.16 (.33)	4.79 (.34E-03)
	January 1968- December 1972	5.56 (.71E-04)	153.17 (0.00)	3.78 (.26E-02)	5.48 (.84E-04)	2.97 (.013)	10.31 (.57E-08)
OWN VARIANCE	January 1973- December 1977	1.57 (.17)	3.24 (.75E-02)	1.42 (.22)	22.89 (0.00)	0.94 (.46)	18.75 (.53E-15)
	January 1978- December 1982	7.36 (.19E-05)	268.89 (0.00)	1.68 (.14)	13.33 (.17E-06)	0.52 (.76)	2.18 (.058)
	AGGREGATE: January 1963- December 1982	102.48 (.45E-12)	2195.00 (0.00)	41.50 (.32E-12)	233.36 (.23E-37)	27.93 (.11)	180.12 (.92E-27)

TABLE 9A: CHI-SQUARED TESTS FOR SIGNIFICANCE OF BASIS PORTFOLIO MEAN RETURNS
(p-values in parentheses)

SAMPLE PERIOD	FIVE FACTORS RAW RETURNS	FIVE FACTORS EXCESS RETURNS	TEN FACTORS RAW RETURNS	TEN FACTORS EXCESS RETURNS	FIFTEEN FACTORS RAW RETURNS	FIFTEEN FACTORS EXCESS RETURNS
January 1963- December 1967	20.53 (.99E-03)	20.23 (.11E-02)	25.98 (.38E-02)	27.28 (.23E-02)	35.48 (.21E-02)	35.52 (.21E-02)
January 1968- December 1972	28.62 (.28E-04)	11.15 (.048)	36.83 (.61E-04)	12.85 (.23)	49.21 (.16E-04)	21.25 (.13)
January 1973- December 1977	15.88 (.82E-02)	3.27 (.66)	23.25 (.99E-02)	10.36 (.41)	27.43 (.025)	12.74 (.62)
January 1978- December 1982	12.96 (.024)	7.56 (.18)	35.23 (.11E-03)	23.35 (.95E-02)	37.48 (.11E-02)	24.58 (.056)
AGGREGATE: January 1963- December 1982	77.69 (.96E-08)	42.20 (.26E-02)	121.29 (.41E-09)	73.83 (.90E-03)	149.60 (.13E-08)	94.10 (.32E-02)

TABLE 9B: ADJUSTED CHI-SQUARED GOODNESS OF FIT TESTS FOR STATISTICAL FACTOR MODELS
(p-values in parentheses)

	FIVE FACTORS	TEN FACTORS	FIFTEEN FACTORS
January 1963- December 1967	.3197E+06 (0.00)	.3083E+06 (0.00)	.3010E+06 (0.00)
January 1968- December 1972	.3200E+06 (0.00)	.3102E+06 (0.00)	.3022E+06 (0.00)
January 1973- December 1977	.3460E+06 (0.00)	.3354E+06 (0.00)	.3271E+06 (0.00)
January 1978- December 1982	.3207E+06 (0.00)	.3087E+06 (0.00)	.3008E+06 (0.00)

TABLE 10: MULTIVARIATE REGRESSION TESTS FOR THE NUMBER OF FACTORS
F-STATISTICS AND AGGREGATE CHI-SQUARED STATISTICS FOR SIGNIFICANCE OF INTERCEPTS
(p-values in parentheses)

SAMPLE PERIOD

NUMBER OF FACTORS UNDER NULL	NUMBER OF FACTORS UNDER ALTERNATIVE	VERSION OF APT	January 1963-December 1967	January 1968-December 1972	January 1973-December 1977	January 1978-December 1982	AGGREGATE: January 1963-December 1982
FIVE	10	ZERO BETA	1.30 (.23)	1.33 (.22)	0.91 (.53)	2.21 (.018)	57.38 (.037)
		RISKLESS RATE	1.21 (.29)	0.54 (.86)	0.87 (.57)	1.53 (.13)	41.45 (.41)
	15	ZERO BETA	1.12 (.33)	1.29 (.21)	0.77 (.71)	1.55 (.090)	71.00 (.16)
		RISKLESS RATE	1.07 (.38)	0.72 (.77)	0.78 (.70)	1.19 (.28)	56.38 (.61)
TEN	5	ZERO BETA	1.61 (.16)	1.23 (.30)	0.44 (.82)	0.38 (.86)	18.31 (.57)
		RISKLESS RATE	1.29 (.27)	0.85 (.51)	0.39 (.86)	0.19 (.97)	13.64 (.85)
	15	ZERO BETA	0.87 (.59)	0.99 (.46)	0.48 (.95)	0.28 (1.00)	39.45 (.98)
		RISKLESS RATE	0.76 (.73)	0.90 (.56)	0.36 (.99)	0.27 (1.00)	34.46 (1.00)
FIFTEEN	5	ZERO BETA	0.78 (.57)	0.42 (.83)	0.23 (.96)	0.30 (.91)	8.66 (.99)
		RISKLESS RATE	0.65 (.66)	0.46 (.81)	0.53 (.75)	0.52 (.76)	10.82 (.95)
	10	ZERO BETA	0.51 (.88)	0.49 (.90)	0.37 (.96)	0.24 (.99)	16.13 (1.00)
		RISKLESS RATE	0.44 (.92)	0.62 (.80)	0.35 (.97)	0.27 (.99)	21.57 (.99)

References

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

BARTLETT, MAURICE S. [1950], "Tests of Significance in Factor Analysis," *British Journal of Mathematical and Statistical Psychology*, Vol. 3, pp. 77-85.

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.

BLUME, MARSHALL E. [1970], "Portfolio Theory: A Step Towards Its Practical Application," *Journal of Business*, Vol. 43, pp. 152-173.

——————[1979], "Stock Returns and Dividend Yields: Some More Evidence," Working Paper 1-79, Rodney L. White Center for Financial Research, University of Pennsylvania.

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.

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 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.

CHEN, NAI-FU, RICHARD W. ROLL AND STEPHEN A. ROSS [1984], "Economic Forces and the Stock Market," Working Paper , Graduate School of Management, UCLA.

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

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 [1983], "Yes, the APT is Testable," Unpublished, School of Management and Organization, Yale University.

ELTON, EDWIN J., MARTIN J. GRUBER AND JOEL RENTZLER [1983], "A Simple Examination of the Empirical Relationship Between Dividend Yield and Deviations from the CAPM," *Journal of Banking and Finance*, Vol. 7, pp. 135-146.

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

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

GEWEKE, JOHN F. AND KENNETH J. SINGLETON [1980], "Interpreting the Likelihood Ratio Statistic in Factor Models When the Sample is Small," *Journal of the American Statistical Association*, Vol. 75, pp. 133-137.

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.

HANSEN, LARS P. AND KENNETH J. SINGLETON [1982], "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, Vol. 50, pp. 1269-1286.

————[1983], "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, Vol. 91, pp. 249-265.

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

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

JORESKOG, KARL G. [1967], "Some Contributions to Maximum Likelihood Factor Analysis," *Psychometrika*, Vol. 32, pp. 443-482.

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 II: The Optimal Construction of Basis Portfolios," Unpublished, Graduate School of Business, Columbia University.

——————[1985b], "Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons," 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.

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

MARSH, TERRY A. [1985], "Asset Model Pricing Specification and the Term Structure Evidence," Working Paper 1612, National Bureau of Economic Research.

MEHRA, RAJNISH AND EDWARD C. PRESCOTT [1985], "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, Vol. 15, pp. 145-161.

MERTON, ROBERT C. [1973], "An Intertemporal Capital Asset Pricing Model," *Econometrica,* Vol. 41, pp. 867-887.

MILLER, MERTON H. AND MYRON S. SCHOLES [1978], "Dividends and Taxes," *Journal of Financial Economics,* Vol. 6, pp. 333-364.

——————[1982], "Dividends and Taxes: Some Empirical Evidence," *Journal of Political Economy,* Vol. 90, pp. 1118-1141.

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

PFLEIDERER, PAUL C. [1983], "A Short Note on the Similarities and the Differences Between the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT)," Unpublished, Graduate School of Business, Stanford University.

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

——————[1981b], "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.

——————[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.

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.

RUBIN, DONALD B. AND DOROTHY T. THAYER [1982], "EM Algorithms for ML Factor Analysis," *Psychometrika*, Vol. 57, pp. 69-76.

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

——————[1983], "Factor Pricing Models: Synthesis and Extensions," Unpublished, Graduate School of Business Administration, University of California Berkeley.

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

——————[1985], "Multi-Beta CAPM Or Equilibrium-APT? A Reply to Dybvig and Ross," Unpublished, 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.

——————[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.