A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm
Author(s): Stephen J. Brown and Mark I. Weinstein
Source: *The Journal of Finance*, Vol. 38, No. 3 (Jun., 1983), pp. 711-743
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2328078
Accessed: 27/11/2009 08:02

THE JOURNAL OF FINANCE • VOL. XXXVIII, NO. 3 • JUNE 1983

A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm

STEPHEN J. BROWN and MARK I. WEINSTEIN*

ABSTRACT

We propose a new approach to estimating and testing asset pricing models in the context of a bilinear paradigm introduced by Kruskal [18]. This approach is both simple and at the same time quite general. As an illustration we apply it to the special case of the arbitrage pricing model where the number of factors is pre-specified. The data appear to be generally in conflict with a five or seven factor representation of the model used by Roll and Ross [30]. When we consider the number of replications of our test and the large number of observations on which it is performed, the frequency with which we reject the three factor APM does not lead us to conclude that this model is unrepresentative of security returns. Further, the rejection of the five and seven factor versions is to be expected if the three factor version is correct. The paradigm gives insight into the appropriate specification of the model and suggests that there may be a small number of economy wide factors that affect security returns.

TESTS OF ASSET PRICING models on the basis of observed stock market data have represented a staple of the financial economics literature of the past twenty years. The early work is reviewed in the famous paper by Jensen [12], and papers by Fama and MacBeth [6]. Rosenberg and Marathe [31], and Roll and Ross [30]. There has been a great deal of confusion regarding what testable propositions are implied by these models (see discussion in Roll [29]). In this paper we show that at a minimum all asset pricing models yet proposed imply a simple easily testable hypothesis which we refer to as the bilinear hypothesis. While specific models may imply further testable propositions, if the bilinear hypothesis is rejected the model is inconsistent with the data and should be rejected.

Our test exploits the fact that the economic content of any equilibrium asset pricing model can be expressed in the form of a potentially refutable constraint on the set of parameters of the process generating security returns. This constraint specifies that, given a set of security specific characteristics, there exists a linear relationship common across securities relating expected returns to these

* Bell Laboratories, Murray Hill, New Jersey. The second author is affiliated with the University of Southern California, Los Angeles, CA 90089. This paper was written while the second author was visiting at Columbia and New York Universities on leave from State University of New York at Buffalo.

The views expressed are those of the authors and do not necessarily reflect those of Bell Laboratories or of the Bell System. This paper has gained considerably from suggestions made in seminars too numerous to mention in three countries on two continents and from the comments of colleagues at the various places we are or have been affiliated with. We would like to thank especially the Editor, Michael Brennan, and referee, Richard Roll, as well as Jack Lee for their extensive comments on the previous draft. Of course, we do not implicate them in the errors that remain.

characteristics. The coefficients of this linear relationship represent the market prices of the different characteristics. In the case of the Sharpe [34], Lintner [23], and Mossin [25] Capital Asset Pricing Model, the characteristic specific to each security is defined in terms of the covariance of the returns on the security with that of a market portfolio, whereas the skewness preference model of Kraus and Litzenberger [17] characteristics are defined in terms of covariance and co-skewness with the market portfolio. The model introduced by Rosenberg and Marathe [31] includes skewness preference and yield preference effects. The recently developed arbitrage pricing model of Ross [33] studied in Roll and Ross [30] differs only in that it does not require that the characteristics be prespecified, though on the other hand it does require that the dimension of the space of characteristics be known in advance.[1] Since expected returns can be expressed as a product of a set of characteristics and a set of market prices attributed to those characteristics, the models share a bilinear representation in that, given the market prices of the characteristics expected returns are linear in the characteristics, and, given the characteristics expected returns are linear in these market prices.

The one constraint that the models share is intrinsic in this bilinear representation. The models require that the market price of each characteristic that is priced is common to all securities. In the simple two parameter market model, the implied risk premium and risk-free rate are constant across securities. This is a fairly weak requirement of the model, but it is one which with few exceptions (Gibbons [8], Hess [11], Stambaugh [37], Brennan [1], and Jobson [13]) has been ignored. Given the bilinear representation of the models, there are, of course, other testable propositions which differ according to the specific model under consideration and the particular definition of risk. The papers by Fama and MacBeth [6], Kraus and Litzenberger [17], and other examine some of these propositions.

In this paper, we shall propose a simple test of the bilinear model constraint, a test that can be applied in the context of many popular asset pricing model variants. For illustrative purposes, we shall then apply the test to the well-known Arbitrage Pricing Model (APM) in the special case where there is assumed a finite prespecified number of factors. This model was examined in the recent paper by Roll and Ross [30] and we use their data.

This example is interesting for a variety of reasons. Not only is it a simple model that is well known in the current literature, but also it is valuable as a case study illustrating the problems and pitfalls associated with testing asset pricing models in a more general context. We find the model somewhat in conflict with the observed data. Since the paradigm we work with includes the familiar two parameter model as well as skewness preference and yield preference generalizations of the model, rejection of our paradigm would have implications that extend beyond consideration of the APM. We examine the sensitivity of our

[1] Of course, in empirical work this dimension could be estimated from observed stock market data. In a recent paper, Shanken [33] argues that the empirical version of this model is consistent with a superfluity of factor representations. We simply note here that this argument does not address the implication that for any given representation, the price of security specific characteristics or factors are constant across groups.

results to changes in the hypothesized number of factors, and to specific industry effects, and find evidence that there may be as few as three economy-wide factors, and certainly no more than five if the APM is correct. We conclude with some observations on the relevance of rejecting the APM on the basis of many observations and low numerical values of the F ratio.

This paper is organized as follows. In Section I, we introduce a bilinear paradigm originally due to Kruskal [18] and show how some common asset pricing models imply constraints on the coefficients of the model. In Section II, we discuss estimation issues in the context of a simple and easy-to-apply test of a special case of the APM. In Section III, we discuss the data used in this study and present the results of the test. Section IV concludes with a discussion of directions for future research. An Appendix presents the details of the test statistic and the relationship between the bilinear representation of the APM and the factor model commonly used to estimate the APM.

I. The Bilinear Paradigm

Perhaps the easiest way to introduce our model is to view it as a generalized representation of the return generating process. For example, consider the return generating process of the two parameter model posited by Fama and MacBeth

$$\tilde{R}_i = \tilde{\gamma}_0 + \tilde{\gamma}_1 \beta_i + \tilde{\eta}_i \quad i = 1, m \tag{1}$$

where, suppressing time subscripts, $\tilde{R}_i$ is the *ex post* return on security i, γ_0 and γ_1 represent the return on a zero beta asset and risk premium respectively, β_i is the measure of systematic risk for security i, and $\tilde{\eta}_i$ is a zero mean stochastic term specific to security i. If we consider t observations on the process described by (1), we have the following matrix equation

$$\underset{txm}{R} = \underset{tx22xm}{AB} + \underset{txm}{U} \tag{2}$$

where

$$\underset{txm}{R} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ R_{t1} & R_{t2} & \cdots & R_{tm} \end{bmatrix}$$

$$\underset{tx2}{A} = [\gamma_0 : \gamma_1], \; B = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_m \end{bmatrix}$$

and the elements of U are zero mean stochastic terms uncorrelated across time but possibly correlated across securities.[2]

Viewed within this framework, most existing tests of asset pricing models are

[2] Kruskal [18] has termed models of the type given by Equation (2) "bilinear." He points out that most models fitted to multivariate data can be represented as special cases of Equation (2). If the matrix A is known, we have the standard multivariate regression model. If the matrix A is not known, then there exists a range of factor analysis type models for examining a given set of data.

tests both of the dimensionality of the A and B matrices and the specific definition of the B matrix. For example, Fama and MacBeth [6] considered (Equation 7)

$$\tilde{R}_i = \tilde{\gamma}_0 + \tilde{\gamma}_1\beta_i + \tilde{\gamma}_2\beta_i^2 + \tilde{\gamma}_3 s_i + \tilde{\eta}_i \tag{3}$$

where s_i represents the standard deviation of residual returns $\tilde{\eta}_i$ for security i; then

$$\underset{tx4}{A} = [\gamma_0 \vdots \gamma_1 \vdots \gamma_2 \vdots \gamma_3]$$

$$\text{and} \qquad B = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_m \\ \beta_1^2 & \beta_2^2 & \cdots & \beta_m^2 \\ s_1 & s_2 & \cdots & s_m \end{bmatrix}$$

The Fama and MacBeth test is a test that in terms of expected values $\gamma_2 = \gamma_3 = 0$ and that γ_0 and γ_1 are equal to specific quantities, namely the risk-free rate and risk premia respectively. Without addressing in detail the procedure used by F&M to estimate the parameters of the model and to test their hypotheses, we suggest there is a simpler test. This test is simpler and yet at the same time more general. It can be applied to the models (1), (3), and to the other models of equilibrium asset pricing that have appeared in the literature.

Consider the bilinear paradigm (2). Partition the m securities into disjoint groups of m_1 and m_2 securities respectively. Then, without loss of generality we can write (2) as

$$[R_1 \vdots R_2] = A[B_1 \vdots B_2] + [U_1 \vdots U_2] \tag{4}$$

where the subscripts refer to the first and second subgroups respectively. Equation (4) can be written

$$[R_1 \vdots R_2] = [AB_1 \vdots AB_2] + [U_1 \vdots U_2] \tag{5}$$

which imposes the bilinear constraint that the market price attributed to security specific characteristics is the same across groups of securities. In other words, in the context of the two parameter model (1), the risk-free rate and risk premia implied for one group of securities should be the same as those implied for another group of securities.

An alternative to (5) is that the bilinearity constraint is violated,

$$[R_1 \vdots R_2] = [A_1B_1 \vdots A_2B_2] + [U_1 \vdots U_2] \tag{6}$$

where[3] $A_1 \neq A_2$. Referring back to the two parameter model (1), in any given sample the risk-free rates and risk premia implied for a given day may well differ

[3] In certain applications to be discussed below, individual elements of A_1 and A_2 may not be identified, in which case the alternative hypothesis is that there does not exist a positive definite matrix G such that $A_1G = A_2$. For example the CAPM can be written as

$$R_i = r_f \cdot 1 + [ER_m - r_f]\beta_i + \eta_i$$

or

$$R_i = r_f[1 - \beta_i] + ER_m\beta_i + \eta_i$$

from one security or group of securities to another. If these implied rates measure the same quantity, one should not expect these differences to be large relative to their error of measurement. On this our test is based.

In principle, one could test the null hypothesis of the bilinear paradigm by maximizing the likelihood function of the observations for the parameters for models (5) and (6) and forming the ratio of maximized likelihoods. For certain regularity conditions (which will be met for most applications of the model) minus two times the logarithm of this ratio will be distributed in large samples as chi square with degrees of freedom equal to the number of restrictions imposed on the alternative model by the bilinearity condition (e.g., Silvey [35], p. 113).

With some further assumptions, it is possible to derive a generalized F ratio that is numerically equivalent to this chi square statistic, in large samples. Such an F ratio can be thought of in the following way. Suppose B and the covariance matrix of the errors are given. Then a test of $A_1 = A_2$ is simply a test of the general linear hypothesis—a Chow test. If the errors are Normally distributed, the resulting F ratio will be distributed as F with $t \times k$ and $t \times (m - 2k)$ degrees of freedom where k is the column dimension of A. If we replace B and the covariance matrix with maximum likelihood estimates of the parameters (cf., Theil [38] p. 402) then the ratio will be approximately F in large samples. In the Appendix we show that with large numbers of observations and securities, this ratio will converge to the likelihood ratio statistic introduced earlier where we assume that errors are Normally distributed and independent across groups.

The idea of testing the bilinear restriction imposed by asset pricing models is not new. Gibbons [8] proposed just such a test,[4] as did Stambaugh [37] and in a

---

where

$$a_{t1} = r_{ft},\ a_{t2} = [ER_{mt} - r_{ft}]$$

or

$$a^*_{t1} = r_{ft},\ a^*_{t2} = ER_m$$

Note that $A = A^*G$ and $B = G^{-1}B^*$ where G is the positive definite matrix

$$G = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

and since $AB = A^*B^*$ the models are not identified without further information. Thus if A_1 differs from A_2 by such a positive definite matrix G, then the alternative model will yield the same observable implications as the bilinear model. Note, however, that for arbitrary A_1 and A_2 there does not exist such a positive definite matrix. Thus the fact that individual elements of A_1 and A_2 in general are not identified does not prevent a test of the bilinearity restrictions. However, this implies that it is not possible to test the APM on the basis of mean returns a point to which we shall refer later.

[4] It is instructive to consider Gibbon's test of the bilinear restriction in terms of the CAPM. Rewriting the model in Equation (1) along the lines suggested in fn. 3,

$$\begin{aligned} R_i &= \gamma_0[1 - \beta_i] + ER_M\beta_i + \eta_i \\ &= \gamma_0[1 - \beta_i] + R_M\beta_i + \varepsilon_i \\ &= a_i + \beta_i R_M + \varepsilon_i \end{aligned}$$

The bilinear restriction implies that γ_0 is constant in the cross-section. Gibbons suggest that one can slightly different context, Cheng and Grauer [3]. The claim made here is that this test is of quite general application and we demonstrate its generality. In the next section, we show that it can be used to test the recent arbitrage pricing model in the special case where the number of factors are prespecified. Not only does the bilinear paradigm provide a simple test of the model, but it also clarifies estimation issues in this context.

II. The Arbitrage Pricing Model

Suppose that individuals agree that the *ex ante* returns generating process can be represented for security i in period τ:

$$\tilde{r}_{\tau i} = E_{\tau i} + \tilde{\delta}_{\tau 1} b_{1i} + \cdots + \tilde{\delta}_{\tau k} b_{ki} + \tilde{e}_{\tau i} \tag{7}$$

where $\tilde{e}_{\tau i}$ is a zero mean finite variance stochastic term, the b_{ji} are a set of weights specific to security i, and $\tilde{\delta}_{\tau j}$ is the jth stochastic factor common across securities independent of the other $(k - 1)$ factors and $\tilde{e}_{\tau i}$. Then Ross [32] demonstrates conditions sufficient that $E_{\tau i}$ be represented[5]

$$E_{\tau i} = \lambda_{\tau 0} + \lambda_{\tau 1} b_{1i} + \cdots + \lambda_{\tau k} b_{ki} \tag{8}$$

where in equilibrium $\lambda_{\tau 0}$ is set equal to the risk-free rate if one exists and $\lambda_{\tau 1}, \ldots, \lambda_{\tau k}$ are interpreted as factor prices specific to period τ.

Consider t observations on the process described by Equations (7) and (8). We may write the matrix of observations of returns

$$\begin{aligned} R &= \lambda_0 \iota' + A_k B_k + U \\ &= AB + U \end{aligned} \tag{9}$$

where

---

both estimate and test this bilinear restriction by imposing that

$$a_i/[I - \beta_i] = a_j/[1 - \beta_j],\ i, j = 1, m$$

By maximizing the likelihood function for this bilinear restriction, one obtains the test of the bilinear restriction given in the text, where in this simple case one can consider each subgroup to consist of a single security. Stambaugh [37] considers a Lagrange Multiplier test in addition to Gibbons's likelihood ratio test of this bilinear CAPM restriction. This test is, of course, asymptotically equivalent to Gibbons's test. In small samples, the tests will differ; Evans and Savin [4] describe the properties of an Edgeworth correction (essentially a degrees of freedom correction) designed to address this.

A slight extension of Gibbons's approach would allow a time subscript on the a_i terms. This procedure would then yield estimates that could be interpreted either as *ex ante* market expectations of $\tilde{\gamma}_{0t}$ and $\tilde{\gamma}_{1t}$ or in Fama's [5] interpretation, realized returns on two portfolios, returns that of course must be the same across securities at a given point of time. This slight extension would correspond to the case where each row of A (in our notation) is different, allowing for a degree of nonstationarity in the model. In the empirical results reported later in the paper we in fact allow rows of A to differ.

[5] Strictly speaking, Ross shows that the difference between the left and right sides of Equation (8) approaches zero as the number of securities under consideration becomes large.

$$\underset{t\times(k+1)}{A} = [\lambda_0 \mid A_k], \qquad \underset{(k+1)\times m}{B} = \begin{bmatrix} \iota' \\ \cdots\cdots \\ B_k \end{bmatrix}$$

$$\text{for} \quad \underset{t\times m}{R} = \{r_{\tau i}\}, \qquad \underset{t\times k}{A_k} = \{\lambda_{\tau j} + \delta_{\tau j}\}$$

$$\underset{k\times m}{B_k} = \{b_{ji}\}, \qquad \underset{t\times m}{U} = \{e_{\tau i}\}$$

$$\text{and} \quad \lambda_0 = \{\lambda_{\tau 0}\}$$

where ι is a column vector of ones.

Kruskal [18] refers to "direct" and "indirect" procedures for estimating the bilinear models of the type indicated by Equation (9). Indirect procedures involve making certain assumptions about A, estimating B, and in a second step using those estimates as if they were the true parameter values, estimating A. In the finance context, a market model representation is often used to estimate the unknown parameters in B using time series data, and in a second stage cross-section data is used to estimate A. This in essence was the Fama and MacBeth [6] procedure. There has been some concern in this literature about the fact that elements of B are thereby measured with error (e.g., Miller and Scholes [24]). The only "direct" procedure–to estimate A and B simultaneously–in a CAPM context we know of is that of Gibbons [8].[6]

If in the arbitrage pricing model context we assume that the rows of U, the idiosyncratic components of returns, are distributed as multivariate normal with mean vector 0, and diagonal covariance matrix D,[7] when the likelihood function for the observations generated by the bilinear model (3) can be written

$$\begin{aligned} L[A, B, D \mid R] &= p[R \mid A, B, D] \\ &= c\cdot |D|^{-t/2}\exp\{-\tfrac{1}{2}\,\mathrm{tr}\, D^{-1}[R - \lambda_0\iota' - A_kB_k]'[R - \lambda_0\iota' - A_kB_k]\} \end{aligned} \quad (10)$$

where c is a constant of proportionality. It is possible to show that the necessary conditions for a maximum of the logarithm of the likelihood function (10) imply the set of simultaneous equations:[8]

$$\begin{aligned} \hat{B}_k &= [\hat{A}_k'\hat{A}_k]^{-1}\hat{A}_k'(R - \hat{\lambda}_0\iota') \\ \hat{A} &= R\hat{D}^{-1}\hat{B}'[\hat{B}\hat{D}^{-1}\hat{B}']^{-1} \\ \text{and } \hat{d}_{ii} &= \hat{s}_{ii}/t \end{aligned} \quad (11)$$

[6] Gibbons's procedure was subsequently applied by Stambaugh [37] and Hess [11]. For details of the procedure, see fn. 4.

[7] For ease of exposition we omit the simple generalization to a general covariance matrix Σ. For some applications of the model we may know in advance that certain factors are not priced in the APM context. In this event, the nonpriced factors would generate a nondiagonal covariance matrix of the rows of U. While this would increase the complexity of estimation, it does not otherwise change the nature of the test. The Appendix describes a simulation experiment designed to assess how sensitive the test is to the Normal distribution assumption.

[8] Clearly these are only necessary conditions. As noted before, there is an identification problem

where $\hat{D}_{ii}$ is the ith diagonal element of $\hat{D}$ and $\hat{s}_{ii}$ is the corresponding element of the matrix $[R - \hat{A}\hat{B}]'[R - \hat{A}\hat{B}]$, and where the hats denote maximum likelihood estimates.[9]

This set of simultaneous equations in the maximum likelihood estimates has the following regression interpretation. Refer back to the first line of Equation (9). Suppose we knew the $(t \times k)$ matrix A_k. Then the first line of Equation (11) can be interpreted as a set of OLS regression equations where $R - \lambda_0 \iota$ is regressed against A_k one column at a time. In other words, the first line of Equation (11) represents m time series regressions of implied excess returns of individual securities on factor prices and factor realizations given by A_k. In a similar way, the second line of Equation (11) represents t cross-section GLS regressions, with heteroskedasticity correction given by the diagonal error covariance matrix D. This covariance matrix can in turn be estimated from the sum of squared errors given by the third line. It would appear natural to replace unknown values of A_k, λ_0, B_k, and D with maximum likelihood estimators, and it is no accident that the resulting simultaneous system yields the necessary conditions for the maximum of the logarithm of the likelihood function Equation (10).

Inspection of these necessary conditions would suggest the following estimation procedure. Take some initial value for B and D and do a cross-section regression period by period incorporating the heteroskedasticity correction implied by the diagonal error variance matrix D to obtain an initial estimate $\hat{A}$. Then revise $\hat{B}$ in time series regressions, one for each security. At that point we re-estimate D and with the new estimates $\hat{B}$ and $\hat{D}$ repeat the process until convergence.[10] This process is not guaranteed to converge for arbitrary initial values of B and D and neither is it likely to be numerically stable. However, if the initial values of B and D are close to the full maximum likelihood solution, or if one incorporates some more structure into the problem, for example, by identifying the factors, convergence should be rapid.

This is precisely the estimation procedure employed by Roll and Ross [29], henceforth RR, a major study in this area. However, instead of proceeding to convergence they stop after one step. As initial estimates of B and D, they employ a maximum likelihood factor analysis procedure due to Jöreskog [15]. In the Appendix we show conditions sufficient for Jöreskog's algorithm to yield a full maximum likelihood estimate of B and D. These conditions do not require that the factors have *ex ante* normal probability distributions, though they do require that expected returns are stationary in time. In order to estimate factors for the universe of securities, Chen [2] uses a slightly modified version of this procedure where a portfolio based estimate of B is used as the starting value and the algorithm proceeds through two steps.[11] In the remainder of the paper, we assume

in that for any nonsingular matrix G, if A, B attain a maximum of the logarithm of the likelihood (4), so will A^* and B^* where $A^* = AG^{-1}$ and $B^* = GB$. This is not a particular difficulty in this context, since it suffices for us that the product AB is identified.

[9] Note that $\tilde{A}_k$ and $\tilde{B}_k$ are defined in terms of the partitions $\tilde{A} = [\hat{\lambda}_0 \mid \tilde{A}_k]$, and $\tilde{B} = \left[\frac{\iota}{\beta_k}\right]$.

[10] Gibbons [8] suggests just such an approach to maximum likelihood estimation in the bilinear representation of the CAPM.

[11] Chen [2] uses maximum likelihood factor analysis (MLFA) to obtain an initial estimate of the matrix B using 180 securities. These securities are grouped into five disjoint equally-weighted

the appropriate conditions are met, so that we may use Jöreskog's algorithm to obtain maximum likelihood estimates of B and D, proceeding in one step via Equation (11) to obtain a maximum likelihood estimate of A.

We should note in passing that as Gibbons [8] points out in the CAPM context, direct estimation of the parameters A and B of the bilinear model avoids entirely the measurement error problem addressed by Miller and Scholes [24] among others.

Given maximum likelihood estimates of the parameters of the model, one can then construct the tests of the bilinear constraint that are described in the previous section. As noted there and in the Appendix, such tests have an intuitive asymptotic F test interpretation. The next section will describe the application of the test to New York and American Stock Exchange data and present empirical results.

III. A Test of the Arbitrage Pricing Model

If we hold that the bilinear model (9) in fact generates observed returns in the context of an arbitrage pricing model with a prespecified (k) number of factors, then as indicated in Section I a simple test of that theory is to check whether the factor structure, factor prices, and implied risk-free rate given by the matrix A are constant across groups of securities at each point of time. Such a test is, of course, a simultaneous test of the model and a particular factor representation of the model.

A. Data

In this analysis of the arbitrage pricing model (APM) we want to compare our results with those results others have obtained. Since RR is among the most widely known studies of the APM, and furthermore relies on maximum likelihood

---

portfolios using a linear programming procedure designed to maximize the difference between groups. To interpret his results in the present context, using deviations from means (thus neglecting the constant term in Eq. 11) the first set of equations (11) can be written

$$\hat{B} = [\hat{A}_k'\hat{A}_k]^{-1}\hat{A}_k'R$$

and substituting in $A_k'R$ from the second set of equations in (11)

$$\hat{B} = \hat{\Omega}^{-1}[\hat{B}\hat{D}^{-1}\hat{B}']^{-1}\hat{B}\hat{D}^{-1}\hat{\Sigma}$$

where $\hat{\Sigma} = \dfrac{1}{t}R'R$. For MLFA $\hat{B}$ one can show $\Omega = \dfrac{1}{t}\hat{A}'\hat{A} = I + [\hat{B}\hat{D}^{-1}\hat{B}']^{-1}$, the variance of the factors plus the sampling variance of factor estimates $\hat{A}$ given $\hat{B}$. This equation for $\hat{B}$ is then used to obtain the revised $\hat{B}$. However, in using measurement error arguments to reduce the number of securities to five, for five factors Chen guarantees a perfect fit in the regression equations defining B. In other words, B is nonsingular. Thus one can obtain

$$\hat{B} = \Omega^{-1}\hat{B}'^{-1}\hat{D}\hat{B}^{-1}\hat{B}\hat{D}^{-1}\hat{\Sigma} = [\hat{B}'\Omega]^{-1}\hat{\Sigma}$$

which is precisely the equation used by Chen. One may legitimately ask the question whether the advantage (other than numerical convenience) outweigh the loss of information going from 180 securities to only five portfolios, instead of to some larger number of disjoint portfolio groupings.

Table I

Data Description

Source:	Center for Research in Security Prices Graduate School of Business University of Chicago Daily Returns File
Selection Criterion:	By alphabetical order into groups of 30 individual securities from those listed on the New York or American Exchanges on *both* 3 July 1962 and 31 December 1972.
Basic Data Unit:	Return adjusted for all capital changes and including dividends, if any, between adjacent trading days; i.e., $[(p_{jt} + d_{jt})/p_{jt-1}] - 1$, where p = price, d = dividend, j = security index, t = trading day index.
Maximum Sample Size per Security:	2619 daily returns
Number of Selected Securities	1260, (42 groups of 30 each)

methods, we have chosen this work as our starting point. In their work, RR make a number of arbitrary decisions regarding data source time period and group size for their initial maximum likelihood factor analyses. While we might have made different decisions if we were starting *de novo*, we cannot think of any compelling reason to change the decisions that RR made. Thus we have, to the extent possible, replicated the RR sample using the 1980 version of the CRSP Daily Returns file.

The data are briefly described in Table I which follows Table I of RR. We start with 42 groups of 30 stocks each arranged alphabetically. We use daily returns (including dividends) for the period from 3 July 1962 to 31 December 1972. In order to be eligible for inclusion in a given group, the common stock must have been listed on one of the two major stock exchanges on both the starting and ending dates for the period.

The methodology that we shall use in our examination of the APM requires that we conduct factor analyses of each group. We combine two consecutive RR groups together to form each of our 60 security groups. We thus obtain 21 groups of 60 securities each. Note, however, that the results obtained by our factor analyses of any of the 30 security groups will not be identical with those reported by RR for two reasons. First, our results differ because our requirement that we be able to perform an analysis for a 60 security grouping means that we must have common observations across 60 securities, as opposed to the 30 securities that were required by RR. Thus in general each of our runs used fewer observations than those of RR. Second, the results differ because we used a different version of the CRSP tape than did RR.[12]

B. Empirical Results

In this section, we report the results of carrying out the test described in Section I and the Appendix on the data described in Section III.A. Because our

[12] Another reason for difference is that we did not use the EFAP program due to Jöreskog that RR used. For our initial factor analysis we used the IMSL subroutine OFCOMM, which is an implementation of the Jöreskog algorithm used in EFAP.

tests depend on the assumed factor model for the observations, we examined the sensitivity of our results to the assumed number of factors and replicated our tests for three, five, and seven factors using the twenty-one groups of sixty securities. These groups were then subdivided into thirty securities for the purpose of testing for the constancy of the A matrix across groups. We examined the sensitivity to industry effects by replicating the tests with the groups of sixty securities chosen to fall in the same general industrial classification and also with each subgroup of thirty securities chosen from an industrial classification that differed from one subgroup to the next. This data selection procedure differs from the alphabetical group selection procedure used by RR. Finally, in results not reported here, we examined the sensitivity of the results to serial dependence in daily returns and to the nonnormal probability distributions of day-to-day returns.[13]

The tests proposed in Section I can be used to test both the factor process (7) as well as the arbitrage pricing model (8). Define

$$r^*_{\tau i} = r_{\tau i} - E_{\tau i}$$

If the matrix Δ_k represents realizations of the (k) factors, $\Delta_k = \{\delta_{rj}\}$, then from Equation (7)

$$R^* = \Delta_k B_k \tag{12}$$

a bilinear model. If we assume[14] that the *ex ante* means of the factors exist and equal zero, and we further assume that the process is stationary so that

$$E_{\tau i} = E_i \forall_\tau$$

Then R^* has the interpretation of being differences from mean returns. Thus, since Equation (12) is a bilinear model, with the additional assumption of stationarity of expected returns we may test the factor process as well as testing the APM depicted in Equation (9).[15]

If the return generating process is indeed given by Equation (7), then the maximum likelihood estimates of constrained error variances $\hat{D}_c$ are obtained from the idiosyncratic variance estimates provided by maximum likelihood factor

[13] These results, which did not materially affect our conclusions, did show that the results were more sensitive to a lack of independence in day-to-day security returns, studied using every fifth observation, than to the nonnormal probability distribution of daily returns, which was examined on the basis of continuously compounded five day returns. As pointed out in Section II, our tests do not require that rates of return have a Normal probability distribution although they do require that the idiosyncratic components of returns are Normal (Cf. fn. 7). The fact that our results are not sensitive to the use of weekly rates of return, addresses at least in part the possible objection that daily rates of return are subject to errors of measurement due to incomplete trading.

[14] Note that the APM does not in fact require that the *ex ante* distribution of the factors have any moments.

[15] The test of the arbitrage pricing model that we propose based on Equation (9) is in fact a joint test of the constancy of factors and of factor prices. If the first test based on Equation (12) were independent of the second test, then failure to reject on the first test and rejection on the second (with appropriate adjustment of levels of significance) would imply that the factor prices differ across groups. In fact, as we shall see, the gross violations of the null hypothesis on the second test we associated with rejection on the first test, so as to render a verdict of "not proven" for the APM applied to those groups (Cf. fn. 1).

Table II

Test of Hypothesis that the Same Three Factors Generate Security Returns

Group	Roll Ross Groups[a]	Chi-square Test for Adequacy of 3 Factors for 60 Security Groups[b]	Chi-square Test for Adequacy of 3 Factors for First RR Group[c]	Chi-square Test for Adequacy of 3 Factors for Second RR Group[d]	F Test for Constancy of Factors Across Groups[e]	Degrees of Freedom[f]	Number of Periods	Number of Securities[g]
1	1, 2	2059	390.6	431.9	1427.7	7209	2403	60
		.000	.057	.001		129762		
2	3, 4	1931	360.3	394.1	.066	6588	2196	60
		.000	.313	.044		118584		
3	5, 6	1861	382.7	412.0	6083.8	4683	1561	60
		.000	.096	.010		84294		
4	7, 8	1991	426.9	455.2	.096	6717	2239	60
		.000	.002	.000		120906		
5	9, 10	1806	345.2	433.0	.133	4935	1645	60
		.000	.532	.001		88830		
6	11, 12	1653	374.8	315.9	.065	3495	1165	59
		.020	.154	.584		61745		
7	13, 14	1850	368.2	389.6	.046	5049	1683	60
		.000	.219	.061		90882		
8	15, 16	1971	422.5	387.9	543.2	7455	2485	60
		.000	.004	.069		134190		
9	17, 18	1764	308.7	350.8	.000	4260	1420	60
		.002	.936	.448		76680		
10	19, 20	2226.9	417.0	407.5	.152	7335	2445	60
		.000	.006	.015		132030		
11	21, 22	2089	467.6	448.0	.102	7098	2366	60
		.000	.000	.000		127764		
12	23, 24	1960	381.9	421.3	.046	7338	2446	60
		.000	.102	.004		132084		
13	25, 26	2009	431.6	352.4	.117	7377	2459	60
		.000	.001	.424		132786		

[a] These groups, initially of thirty securities each, were chosen according to the alphabetical order of securities on the CRSP tapes of December 1980. This was the procedure employed by Roll and Ross [29]. However, the groups do not exactly match those of Roll and Ross due to the fact that the alphabetical order of securities differed from that of the earlier version of the CRSP tapes used by RR, and also the fact that securities whose returns were perfectly explained by the statistical representation of the factors (the so-called Heywood case) were excluded from the analysis.

[b] χ^2 statistic for test of adequacy of the factor representation across 60 securities. The p value given below the statistic is the probability that 3 factors are sufficient.

[c] χ^2 statistic for test of adequacy of the factor representation across the first Roll-Ross Group. The p value given below the statistic is the probability that 3 factors are sufficient.

[d] χ^2 statistic for test of adequacy of the factor representation across the second Roll-Ross Group. The p value given below the statistic is the probability that 3 factors are sufficient.

[e] F-test of constraint that the same given set of factors generate returns for all 60 securities. p values are omitted because any F marginally greater than 1 is significant. This F statistic is defined in the Appendix, Equation (A.8).

[f] (numerator/denominator) degrees of freedom for the F statistic.

[g] The total number of securities after exclusion of Heywood case securities (see fn. a above).

Table II—*Continued*

Group	Roll Ross Groups[a]	Chi-square Test for Adequacy of 3 Factors for 60 Security Groups[b]	Chi-square Test for Adequacy of 3 Factors for First RR Group[c]	Chi-square Test for Adequacy of 3 Factors for Second RR Group[d]	F Test for Constancy of Factors Across Groups[e]	Degrees of Freedom[f]	Number of Periods	Number of Securities[g]
14	27, 28	1886	377.6	367.4	.061	7200	2400	60
		.000	.132	.227		124800		
15	29, 30	2034	418.7	408.3	.000	7359	2453	60
		.000	.005	.014		132462		
16	31, 32	1826	391.3	380.1	.168	7077	2359	60
		.000	.054	.114		127386		
17	33, 34	1825	417.4	382.7	.105	6966	2322	60
		.000	.006	.097		125388		
18	35, 36	1835	366.6	377.9	.284	4548	1516	60
		.000	.236	.130		81864		
19	37, 38	1776	361.0	352.6	.059	5865	1955	60
		.001	.305	.420		105570		
20	39, 40	2210	442.9	489.6	7.4×10^5	5544	1848	60
		.000	.000	.000		99792		
21	41, 42	1869	435.7	429.9	.083	7302	2434	60
		.000	.001	.002		131436		

analyses of sixty securities taken together, and the unconstrained error variances $\hat{D}_u$ are obtained from factor analyses of the two groups of thirty securities taken separately. Using the estimates of constrained and unconstrained error variances, we may compute the Chi-square and F statistics, in both cases rejecting the null hypothesis of consistency when these statistics are large relative to their critical values. Note that rejecting the consistency of factors across groups is not necessarily rejecting the APM; it merely rejects the hypothesis that a given factor model is descriptive of the return generating process for the universe of securities. Either the model is inappropriate or more than the given number of factors are required to adequately describe stock price changes.

Table II presents the results of these tests where the same three factors are assumed to generate observed security returns. The first fact to note is that for none of the sixty security groups do the Chi-square tests indicate that three factors are sufficient at the commonly used fifty percent level of significance.[16]

In group 6 we have the interesting phenomenon that at least one security return was perfectly correlated with the factors, the so-called "Heywood case." While this may be a mere statistical artifact, it is consistent with that security

[16] This is by no means conclusive evidence of model failure, however. The Chi-square statistic used in this particular test, a test also used by RR, requires that the common factors follow an *ex ante* Normal probability distribution. Since the test of factor constancy across groups does not require such an assumption, more weight should be given those results, given the accumulation of evidence that would suggest that daily stock returns are non-Normal (e.g. RR, p. 1095). For the sensitivity of the Chi-square test to non-Normal data generating processes, see Olsson [26] and Sinclair [36].

representing a single industry factor,[17] a problem that became more evident as the number of securities increased. In other words, the evidence may be insufficient to distinguish the industry and idiosyncratic components of returns. For this reason the security was excluded from the analysis, and the results on Table II for group 6 apply to the remaining 59 securities.

Table II reports the F test briefly described in Section I and derived in the Appendix (Equation (A8)). There is no need to report significance as every F statistic marginally greater than 1.0 had a p value of 1.0 for the reported degrees of freedom. The hypothesis that the same three factors generate returns is rejected only four times. We note in passing that with over 100,000 observations small differences can magnify into large F ratios! It would seem that while more than three factors would appear necessary to yield a satisfactory statistical representation of the return generating process, the three factors that best represent the observed variation in the data do not significantly differ across groups.

The APM imposes two restrictions on the matrix of factor prices:
(1) the risk-free rate should be constant across groups; and
(2) the factor prices should be constant across groups at a given point in time.
We can construct tests of each of these hypotheses separately as well as the joint test of both hypotheses by varying the restrictions imposed on the A matrix.[18]

We could consider tests of hypotheses (1) and (2) in either of two ways. Either we could assert that (1) and (2) hold on a day-by-day basis or that the equalities hold on the basis of the time-series-mean factor prices. On the one hand, the APM is a one period model which constraints *ex ante* expectations to be constant across groups of securities. On any given period, this hypothesis can be tested on the basis of cross-section data using an asymptotic F test or Chi-square test. Are estimated factor prices at each time significantly different across groups? On the other hand, one could make the additional assumption that these *ex ante* expectations are the same through time and thereby obtain a more powerful test using a single cross-section regression of mean security returns. Use of this additional information will (if correct) lead to more precise estimates of the within group *ex ante* expectations, and in many instances it will be perfectly acceptable to base the tests on sample means (as well as being easier computationally).

Unfortunately, one cannot apply our bilinear test to the APM on the basis of mean returns unless the factors are identified. Suppose we have three factors and we estimate the vector $\mathbf{a}'$ on the basis of regressing sample mean returns

[17] The security that was excluded was Day Mines, Inc., a small mining venture that was included in our sub-group 12 along with Dome Mines Limited, a much larger mining corporation. The problem was evident though not as severe for this security when five, seven, and ten factors were considered in the analysis. Day Mines experienced the highest standard deviation of returns, .031, ten percent higher than that of Dennison Manufacturing, the next highest in the group and half again as high as that of Dome Mines, while the covariation of returns was higher between the two mining companies than between any other two securities in the group. Given the high variability of its returns, Day Mines, Inc. was in effect chosen by the algorithm to represent a mining and minerals factor.

[18] As noted in fn. 1 and fn. 15, since this test is in reality a joint test of the APM and the factor generating process, we are implicitly assuming here that, with the possible exceptions of groups 1, 3, 8, and 20, the three factor process is descriptively accurate.

on factor loading matrices $\underset{3\times m_1}{\overset{1\times 3}{B_1}}$ and $\underset{3\times m_2}{B_2}$. Suppose $\mathbf{a}'_1$ differs from $\mathbf{a}'_2$ so estimated. Then there always exists[19] a rotation $\underset{3\times 3}{G}$ of B_2, $\underset{3\times m_2}{B_2^*} = \underset{3\times 3}{G}\ \underset{3\times m_2}{B_2}$ such that $\mathbf{a}_2^{*\prime} = \mathbf{a}'_2G^{-1} = \mathbf{a}'_1$. If we know $\mathbf{a}'_2$ and B_2 only up to such rotations, then we can never tell $\mathbf{a}'_2$ from $\mathbf{a}_2'^* = \mathbf{a}'_1$, regardless of the particular values of $\mathbf{a}'_2$ and $\mathbf{a}'_1$. Thus it is not possible to apply our test to the APM on the basis of mean returns.

To put this another way: we saw before (fn. 4) that one of the strengths of our residual based tests is that it requires only that the product AB be identified. By this we mean that for arbitrary A_1 and A_2 matrices of factor prices and factor realizations, there does not exist a nonsingular matrix G such that $A_1 = A_2G$. In this instance we need at least four (one greater than the number of factors) cross-section estimates of $\mathbf{a}'_2$ representing the rows of A_2.

This argument does not require that the test be conducted on a day-by-day basis. Subperiod results could be used if one were concerned about the power of the test and wished to exploit the assumption that *ex ante* means are constant through time. Note that this identification problem does not preclude tests on the implied risk-free rate using means, since this quantity is always identified[20] (cf. Jobson [13]).

Indeed, it could be argued that the tests conducted on a day-by-day basis are too stringent in view of the large number of daily factor prices and realizations that must be assumed equal across groups. Some adjustment of the size of the

[19] To see this it might be helpful to consider a simple numerical example. Suppose we estimate

$$\mathbf{a}'_1 = [.01 \quad .05 \quad .10]$$

$$\mathbf{a}'_2 = [.50 \quad .02 \quad .30]$$

where $\mathbf{a}'_2$ is estimated on the basis of

$$B_2 = \begin{bmatrix} -.50 & .30 & .10 & .40 & .10 \\ .50 & -.20 & .20 & .20 & 2.10 \\ .50 & -.10 & -.30 & -.61 & -.04 \end{bmatrix}$$

$$\text{Then } \mathbf{a}_2^{*\prime} = \mathbf{a}'_2\mathrm{G}^{-1} = [.5 \quad .2 \quad .3] \begin{bmatrix} .02 & 0 & 0 \\ 0 & .25 & 0 \\ 0 & 0 & .33 \end{bmatrix}$$

$$= [.01 \quad .05 \quad .10] = \mathbf{a}_1$$

and

$$B_2^* = GB_2 = \begin{bmatrix} 50 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -.50 & .30 & .10 & .40 & .10 \\ .50 & -.20 & .20 & .20 & 2.10 \\ .50 & .10 & -.30 & -.61 & -.04 \end{bmatrix}$$

$$= \begin{bmatrix} -25 & 15 & 5 & 20 & 5 \\ 2 & -.80 & .80 & .80 & 8.40 \\ 1.5 & .9 & -.9 & -1.83 & -.12 \end{bmatrix}$$

[20] It is also possible to use means in cases where the factors are prespecified. For example, in the CAPM it would be perfectly appropriate to construct a cross-section test based on implied risk-free rates and risk premia estimated on the basis of time series mean returns.

tests is appropriate and in Section III.C. we explicitly address one approach to this problem and show how it affects the results that we report.[21]

The tests of the risk-free rate and factor prices are carried out exactly as in the bilinear test of the constancy of factor realizations across groups. Under the null hypothesis that the factor prices and realizations are constant across groups, we use the time series of returns on sixty securities to estimate $B = \{B_1 \mid B_2\}$. Under the hypothesis that they differ, we estimate B_1 and B_2 separately on the basis of thirty securities each. Note that under the null hypothesis these estimates should be the same up to rotation. For the null hypothesis, we estimate A imposing constraints (1) and (2) above by regressing the cross-section of returns on B using generalized least squares with error covariance matrix D estimated on the basis of the idiosyncratic variances from the time series of returns under the alternative hypothesis. The two A matrices for the alternative hypothesis are estimated in the same way. As demonstrated in the Appendix, the likelihood ratio test described in Section I amounts to a comparison of weighted sums of squared residuals from these cross-section regressions.

We consider three tests: the first that the implied risk-free rate is constant across groups but potentially different from one period to the next, the second that the implied factor prices and realizations are constant in the cross-section and the third, that both constraints hold. The results of the three tests on each of the 21 groups of securities are reported in Table III. The column labeled F_1 reports the F-statistic for the hypothesis that the implied risk-free interest rate is the same in both subgroups. F_2 is the statistic for the hypothesis that the factor prices are the same in both subgroups and F_3 tests the joint hypothesis that the risk-free rate and the factor prices are the same in both subgroups. These F statistics are defined in the Appendix Equations (A.13), (A.15) and (A.17) respectively. We do not report significance as all p values are very close to 1.0 for F marginally greater than unity, and close to zero for F marginally less than unity given the number of degrees of freedom reported.[22]

The first hypothesis, that the risk-free rate is constant across subgroups, is rejected 21 out of 21 times. The second hypothesis is rejected in every group, as is the joint hypothesis. Thus, Table III would appear to provide strong evidence that realized security returns are inconsistent with the three-factor representation of the APM.

[21] It has also been suggested that the day-by-day test is too stringent because sampling error from period to period that is uncorrelated across groups may lead to spurious rejections. This period by period variability is captured by the matrix D which in our application of the test is estimated on the basis of the time series of errors from the unrestricted model. It is precisely for this reason that we use the inverse of D to weight residuals in the comparison of restricted and unrestricted cross-section sums of squares and for this reason also that the matrix D^{-1} enters into the definition of the maximum likelihood estimates of parameters of interest. Further, it is clear that the effect, if any, of this measurement error is to introduce more "noise" in our estimates of A and hence reduce the power of our test. Nevertheless, as we shall see, we still reject the five and seven factor versions of the APM even at adjusted significance levels. This is further evidence that measurement errors present no problem in this case.

[22] For those that are concerned about the knife edge properties of tests when there are large numbers of degrees of freedom; note that $F_{.95}(120, 120)$ is only 1.35. We also computed the appropriate likelihood ratio statistics, but as expected they led to the same conclusions as were drawn from an examination of the F ratios.

Table III

Test of Three Factor APM[a]

Group	Test of Equality of Implied Risk-free Rates (F_1)	Degrees of Freedom[b]	Test of Equality of Factor Premia (F_2)	Degrees of Freedom[b]	Test of Equality of Risk-free Rates and Factor Premia (F_3)	Degrees of Freedom[b]
1	1.25	2403 124956	1153.4	7209 124956	1099.5	9612 124956
2	1.12	2196 114192	1.22	6588 114192	1.18	8784 114192
3	1.33	1561 81172	515.48	4683 81172	656.0	6244 81172
4	1.09	2239 116428	1.20	6717 116428	1.17	8956 116428
5	1.14	1645 85540	1.39	4935 84450	1.32	6580 85540
6	1.08	1165 59415	1.15	3495 59415	1 .11	4660 59415
7	1.10	1683 87516	1.24	5049 87516	11.8	6732 87516
8	1.15	2485 129220	193.7	7455 129220	203.7	9940 129220
9	1.06	1420 73840	1.34	4260 73840	1.25	5680 73840
10	1.06	2445 127140	1.23	7355 127140	1.18	9780 127140
11	1.10	2366 123032	1.615	7098 123032	1.48	9464 123032
12	1.06	2446 127192	1.24	7338 127192	1.19	9784 127192
13	1.18	2459 127868	1.36	7377 127868	1.29	9836 127868
14	1.06	2400 124800	1.25	7200 124800	1.18	9600 124800
15	1.15	2453 127556	1.19	7359 127556	1.17	9812 127556
16	1.04	2359 122668	1.57	7077 122668	1.44	9436 122668
17	1.07	2322 120744	1.37	6966	1.31	9288 120744
18	1.16	1516 78832	4.58	4548 78832	3.72	6064 78832
19	1.07	1955 101660	1.83	5865 101660	1.62	7820 101660
20	1.21	1848 96096	8.12	5544 96096	6.26	7392 96096
21	1.14	2434 126568	1.29	7302 126568	1.24	9736 126568

[a] F_1: F ratio for test of constancy of implied risk-free rate across groups. F_2: F ratio for test of constancy of factors and factor prices across groups. F_3: F ratio for test of constancy of both risk-free rate and factors and factor prices. These F ratios are defined in Equations (A.13), (A.15) and (A.17) of the Appendix.

[b] (numerator/denominator) degrees of freedom. Note that p values for all F statistics are approximately 1.0.

However, an interesting pattern emerges from a comparison of Tables II and III. While it is true that we reject the three-factor APM restrictions every time they are tested, we do so on the basis of small numerical values of the F ratio *except* where on Table II we reject the hypothesis that the factors are the same across groups. In Section III.C, we shall argue that the level of significance of the test should reflect the very large numbers of observations and that the small numerical values of the F ratios observed should not be construed as strong evidence against the model.

In view of the fact that Tables II and III are by no means compelling evidence for a three factor model, Table IV compares results for three, five, and seven factor models. The column headed $p(\chi^2) > .5$ gives the number of thirty security subgroups for which the given factor model is an adequate representation of the data, at least according to the Chi-square test, and the columns headed by F_1, F_2 and F_3 correspond to the F statistics reported on Table III. The rows correspond to the F values reported on Table II, which test whether the factors are the same across groups. The most striking result observed on Table IV is that while the adequacy of the given factor model (measured by the Chi-square statistics) increases with the number of factors, the proportion of securities for which the factors are the same across groups *decreases* with the number of factors. However, the conclusions drawn from this table with respect to the APM are similar to the conclusions drawn from Tables II and III; where the factors do not significantly differ across groups, the APM is rejected, but only with small numerical values of the F ratio. For one group of sixty securities, the five factor APM is actually accepted.

In view of the fact that the test we propose is a joint test of the number of factors and the APM, if we accept a particular factor model and our test has any power, we must expect to reject the model for other factor representations. Thus the fact that we reject the three factor model only at very small numerical values of the F ratio, and reject other representations of the model at larger F ratios can be viewed as strong evidence in *favor* of the three factor APM.

When these results are combined with the fact not reported on Table IV that, as the number of factors increases, increasing numbers of securities must be excluded on the basis of excess correlation with the factors, one might tentatively conclude that the increased number of factors does not reflect economy wide effects but rather specific firm or industry factors. Where the factor model is a sufficiently good representation of the data generating process, we find that the expanded factor APM models do no better and sometimes worse than the three factor model in explaining observed returns. It would appear that if the APM is correct, it is the economy wide factors that are being priced. Further evidence that our anomalous results relative to the APM may be due to idiosyncratic or specific industry factors rather than mispricing of economy wide factors can be found on Table V. In the second block of this table, we exclude from the analysis securities that are very highly though not perfectly correlated with the factors (that is, securities with communalities[23] greater than .999), and report results for the five factor APM. While we still in most cases reject the APM at low numerical

[23] In our context communalities are analogous to the R^2 statistic in regression analysis.

Table IV

Tests of Factor Models and APM for 3, 5, and 7 Factors[a]

F Test for Constancy of Factors Across Groups		Chi-Square Test for Significance of Specified Number of Factors		Test of Equality of Implied Riskfree Rates (F_1)			Test of Equality Factor Premia (F_2)			Test of Equality of Riskfree Rates and Factor Premia (F_3)			
		$p(\chi^2) > .5$	Total	$F < 1$	$1 \leqslant F < 20$	$F > 20$	$F < 1$	$1 \leqslant F < 20$	$F > 20$	$F < 1$	$1 \leqslant F < 20$	$F > 20$	Total
3 Factors	$F < 1$	3	36	0	18	0	0	18	0	0	18	0	18
	$F \geqslant 1$	0	6	0	3	0	0	0	3	0	0	3	3
	Total	3	42	0	21	0	0	18	3	0	18	3	21
5 Factors	$F < 1$	15	30	0	15	0	1	14	0	1	14	0	15
	$F \geqslant 1$	6	12	2	4	0	0	0	6	0	0	6	6
	Total	21	42	2	19	0	1	14	6	1	14	6	21
7 Factors	$F < 1$	7	10	0	5	0	0	5	0	0	5	0	5
	$F \geqslant 1$	28	30	1	14	0	0	1	14	0	2	13	15
	Total	35	40	1	19	0	0	6	14	0	7	13	20
Overall	$F < 1$	25	76	0	38	0	1	37	0	1	37	0	38
	$F \geqslant 1$	34	48	3	21	0	0	1	23	0	2	22	24
	Total	59	124	3	59	0	1	38	23	1	39	22	62

[a] The column labeled $p(\chi^2) > .5$ refers to the number of thirty security subgroups with Chi-square values having probability levels greater than .5; F_1, F_2 and F_3 correspond to the F ratios reported on Table III, with the numbers referring to the number of F values in the designated ranges. The rows refer to F tests reported on Table II and defined in Equation (A.8) used for assessing the adequacy of the designated factor model representation. Note that in no case did such an F ratio fall in the interval $1 \leqslant F < 20$. The reason for finding only 40 thirty security groups in the seven factor representation was the lack of convergence for one sixty security group in that instance.

Table V

Effect of Excluding Securities Highly Correlated with Factors on Tests of Five Factor Models and APM[a]

F Test for Constancy of Factors Across Groups		Chi-Square Test for Significance of Specified Number of Factors		Test of Equality of Implied Riskfree Rates (F_1)			Test of Equality Factor Premia (F_2)			Test of Equality of Riskfree Rates and Factor Premia (F_3)			
		$p(\chi^2) > .5$	Total	$F < 1$	$1 \leq F < 20$	$F > 20$	$F < 1$	$1 \leq F < 20$	$F > 20$	$F < 1$	$1 \leq F < 20$	$F > 20$	Total
Alphabetical	$F < 1$	15	30	0	15	0	1	14	0	1	14	0	15
Groups: Excl.	$F \geq 1$	6	12	2	4	0	0	0	6	0	0	6	6
Heywood Cases	Total	21	42	2	19	0	1	14	6	1	14	6	21
Alphabetical	$F < 1$	25	42	1	20	0	4	17	0	4	17	0	21
Groups: Excl.	$F \geq 1$	0	0	0	0	0	0	0	0	0	0	0	0
$\rho^2 > .999$	Total	25	42	1	20	0	4	17	0	4	17	0	21
Industry	$F < 1$	4	18	0	9	0	0	8	0	0	8	10	9
Groups: Excl.	$F \geq 1$	1	6	1	2	0	0	0	3	0	0	3	3
Heywood Cases	Total	5	24	1	11	0	0	8	4	0	8	4	12
Industry	$F < 1$	8	24	1	11	0	1	11	0	1	11	0	12
Groups: Excl.	$F \geq 1$	0	0	0	0	0	0	0	0	0	0	0	0
$\rho^2 > .999$	Total	8	24	1	11	0	1	11	0	1	11	0	12

[a] The columns in this table correspond to those reported for Table IV. The first panel is taken from the five factor results reported on Table IV while the third panel results are further elaborated on in Table VI where the industry groups are defined. Panels 2 and 4 report results excluding securities highly correlated ($\rho^2 > .999$) with the factors.

values of the F ratio, the anomalous F values vanish and we actually accept the model in four instances. This was true for the three and seven factor results as well as for the industry specific results reported in the last two panels of that table and discussed in greater detail below. This analysis also showed that the seven factor model did no better explaining the data than the three or five factor models, suggesting again a relatively small number of economy wide factors.[24]

Table VI contains results comparable to those reported on Table IV for securities chosen according to industrial classifications. There were six two-digit SIC classifications which contained at least sixty companies which had securities listed in 1962 as well as 1972 and where the four digit code was constant over this period.[25] Industry groups one through six comprised sixty securities chosen at random from those industrial classifications, thus ensuring that each pair of thirty security groups came from the same industrial classification. To examine the effect of allowing the industry to differ from one group of thirty securities to the next, the remaining six groups comprise securities chosen thirty at a time from randomly selected industrial classifications with sufficient data for the tests.[26]

Whether we examine the first six industrial groupings or the second six, the results seem broadly similar to the results reported on the previous tables. Whatever factor model assumed for the data, we reject the APM hypothesis at standard levels of significance. With one exception, where the factors do not differ significantly across groups, we reject the APM hypothesis, but at a relatively low numerical value of the F ratio.

However, inspecting Table VI an interesting result emerges. Whatever the factor model assumed for the data, the Chi-square values indicate that the factor model is a *worse* fit for the thirty security industrial groupings than for the alphabetical groupings. The F ratios for the first six groups–where the industry is the same–are slightly but not significantly lower than those for the second six groups where the industrial classifications differ. This result suggests that as we would expect the industry groupings reveal industry factors in addition to economy wide factors evident in the alphabetical groupings which are very much a random selection of industries.

There remains one apparent difficulty. Why is it that the first six industrial groupings are no more successful in their factor representations than the second six, where the industries differ? In part this is due to the rather broad level of aggregation within two digit SIC codes. Code 35, Machinery Except Electrical,

[24] These results are much in the spirit of similar results found by Kryzanowski and To [19]. In their work, increasing the number of securities increased the number of factors, but these additional factors were not stable over time. Sinclair [36] suggests that "blind" factor analysis applied to security data will generally overstate the true number of factors.

[25] These classifications were 20: Food and Kindred Products; 28: Chemicals and Allied Products; 35: Machinery Except Electrical; 36: Electrical and Electronic Machinery Equipment and Supplies; 37: Transportation Equipment; and 49: Electric Gas and Sanitary Services.

[26] The second set of six groups comprised industry codes 37,35; 49,20; 36,28; 35,49; 20,36; and 49, 37. It is true that there is some duplication of securities going from the first to the second set of six groups, but this was unavoidable given the limited number of two digit SIC codes for which there were substantially more than sixty securities with complete data.

Table VI

Test of Factor Models and APM Across Industry Groups[a]

F Test for Constancy of Factors Across Groups		Chi-Square Test for Significance of Specified Number of Factors		Test of Equality of Implied Riskfree Rates (F_1)			Test of Equality Factor Premia (F_2)			Test of Equality of Riskfree Rates and Factor Premia (F_3)			
		$p(\chi^2) > .5$	Total	$F < 1$	$1 \leq F < 20$	$F > 20$	$F < 1$	$1 \leq F < 20$	$F > 20$	$F < 1$	$1 \leq F < 20$	$F > 20$	Total
3 Factors	$F < 1$	0	24	0	12	0	1	11	0	0	12	0	12
	$F \geq 1$	0	0	0	0	0	0	0	0	0	0	0	0
	Total	0	24	0	12	0	1	11	0	0	12	0	12
5 Factors	$F < 1$	4	18	0	9	0	0	8	1	0	8	1	9
	$F \geq 1$	1	6	1	2	0	0	0	3	0	0	3	3
	Total	5	24	1	11	0	0	8	4	0	8	4	12
7 Factors	$F < 1$	2	2	1	0	0	0	1	0	0	1	0	1
	$F \geq 1$	13	20	3	7	0	0	1	9	0	1	9	10
	Total	15	22	4	7	0	0	2	9	0	2	9	11
Industry Same	$F < 1$	5	24	1	11	0	0	12	0	0	12	0	12
	$F \geq 1$	6	10	2	3	0	0	0	5	0	0	5	5
	Total	11	34	3	14	0	0	12	5	0	12	5	17
Industry Different	$F < 1$	1	20	0	10	0	1	8	1	0	9	1	10
	$F \geq 1$	8	16	2	6	0	0	1	7	0	1	7	8
	Total	9	36	2	16	0	1	9	8	0	9	8	18
Overall	$F < 1$	6	44	1	21	0	1	20	1	0	21	1	22
	$F \geq 1$	14	26	4	9	0	0	1	12	0	1	12	13
	Total	20	70	5	30	0	1	24	13	0	22	13	35

[a] The columns in this table correspond to those reported for Tables IV and V. The results for "Industry Same" and "Industry Different" represent an aggregation over 3, 5, and 7 factor models, with "Industry Same" referring to sixty security groups chosen from SIC two digit codes 20, 28, 35, 36, 37, and 39, and "Industry Different" refers to thirty security subgroups drawn from codes (37, 35), (49, 20), (36, 28), (35, 49), (20, 36), and (49, 37). As in Table IV, a different number of groups for the seven factor representation was caused by one lack of convergence for that representation.

covers a wide range of industrial activities: IBM is included in that group, as is John Deere Company. For this reason one should no more expect that the specific industry factors are the same for two groups of thirty securities chosen from SIC code 35, than we should expect the specific industry factors are the same across SIC codes 35 and 37, Transportation Equipment. Further inspection of the results for groups drawn from code 35 reveals the syndrome reported earlier: the return for one security, Monarch Machine Tool, is highly correlated with the factors. This security was that of a corporation whose activities were highly related to that of another security in the same industrial classification, Cincinnati Milacron, but which experienced a much greater variability of return. Apparently, the factor analysis algorithm identified this security's return as some sort of industry factor. The similarity of results for the groups organized by industrial classification is yet further evidence that the number of economy wide factors are small; and the remaining factors are specific firm or industry effects that may be diversified and not priced in an APM scenario.

C. On the Appropriate Significance Level of the Test

It is tempting to conclude that this evidence somehow represents an overwhelming rejection of the APM; after all, we reject the hypothesis of equality factor prices almost every time we test it. However, our results depend crucially on the adequacy of the given factor model of the return generating process, the normality of the idiosyncratic components of returns, and the effect of extreme numbers of observations on tests of the respective hypotheses.

This latter point is often overlooked in the empirical finance literature. We report results based typically on 120,000 security prices in each group. Given this large number of observations, virtually any hypothesis other than the purely artifactual will be rejected at standard levels of statistical significance. Holding the Type I error constant there is a large decrease in Type II error as the sample size increases, so that the relative importance of the Type I error actually increases.[27]

Not even the most diehard advocate of the APM would state that model was exactly and literally true both in its structure and in the stochastic specification necessary to estimate it.[28] In this sense, to reject the null hypothesis on the basis of many observations is to reject a straw man.

In a classical statistics context, one resolution of the problem is to choose the level of significance in light of the number used to estimate the model. A more satisfactory resolution in our view is to take a Bayesian approach and allow degrees of belief in the respective models to be modified by the data, and compute

[27] This paradox of large numbers attributed to Lindley [22] is well-known in the Bayesian literature (see Leamer [21]). The distinction is often drawn between statistical and economic significance. Since, in this context we only know A_1 and A_2 up to scale and rotation, it is difficult to assign economic significance to $A_1 \neq A_2$. In addition, it was suggested before that day-by-day tests of the APM are too stringent and that some adjustment to the size of the test would appear necessary. We provide here a more formal justification of this intuition.

[28] For a discussion of some of the contradictions that may be implied by such an extreme view, see Shanken [33].

a posterior odds ratio as

$$O_{12} = P[M_1 \mid D]/P[M_2 \mid D]$$

where $P[M_i \mid D]$ is the probability that model i is correct given a set of data D.[29] If one starts with diffuse priors across the models and the parameters of each model, Klein and Brown [16] show that under fairly general conditions, for comparison of two linear models with normally distributed errors,

$$O_{12} = T^{-d/2}(S_2^2/S_1^2)^{T/2}$$

where T represents the number of observations, S_1^2 and S_2^2 represent the sum of squared errors for the two models respectively, and d is the difference in dimensionality between the models.[30] Note that this quantity can be computed from the output of most standard econometric packages.

This statistic bears an obvious relationship to an F statistic for testing a general linear hypothesis, one which takes into account the number of observations exponentially weighted by the difference in dimensionality between the models. Thus, one approach towards choosing a critical value of F would be to choose that value corresponding to a posterior odds ratio of one. For some selected numbers of degrees of freedom, critical F values are shown in Table VII.[31] For example, if we have 10,000 and 100,000 degrees of freedom, a typical case reported in the tables, the critical value of the statistic changes from 1 to (roughly) 16. This is quite a difference and may lead to changed inferences.

The difference between two critical values becomes especially important when we consider the hypothesis that factors and factor prices are the same across groups. Whereas a traditional test rejects this complete set of hypotheses almost every time, the adjusted significance level rejects the APM only half of the time and then only when the given factor model appears to be a bad representation of the original data. In particular, when we consider the number of replications of our test, the frequency with which we reject the three factor APM does not lead us to conclude that this model is unrepresentative of security returns.

IV. Conclusion

In this paper we propose a general approach to the testing of equilibrium asset pricing models. In principle, this approach can be used to test the common bilinear implication of all such models, as is simple and easy to apply.

We have chosen to illustrate its application in the context of the arbitrage pricing model recently introduced by Ross [32] and empirically examined by Roll

[29] For a more complete discussion of the posterior odds criterion, see Zellner [39] and Leamer [21].

[30] Leamer [21] also suggests this definition of the odds ratio. Implicit in his analysis is the requirement that among other things there exists a natural unit of measurement for the independent variables. This condition is violated in the present application. However, Klein and Brown [16] show that this expression follows without such a requirement by directly minimizing measures of prior information specific to the models in question.

[31] In this application of the posterior odds criterion, for simplicity we assume that B is given, replacing it in the definition of the test statistic with the maximum likelihood estimate B. This is not entirely satisfactory, but the approximation should not make a significant difference in the context of the large samples we report in the paper.

Table VII

Values of F_{n_1,n_2} such that the Posterior Odds Ratio is Unity[a]

n_2/n_1	10	50	100	500	1000	5000	10000	15000	20000
100	4.90	5.52	5.69	4.62	3.72	1.90	1.37	1.13	.98
500	6.39	7.04	7.75	10.45	10.93	8.12	6.29	5.32	4.69
1000	7.03	7.50	8.06	11.37	13.42	13.54	11.33	9.87	8.84
5000	8.56	8.73	8.94	10.65	12.71	23.66	28.22	29.20	28.99
10000	9.24	9.34	9.47	10.53	11.87	21.78	30.07	34.79	37.39
50000	10.83	10.86	10.90	11.21	11.61	15.01	19.61	24.35	29.04
100000	11.52	11.54	11.56	11.74	11.97	13.90	16.50	19.27	22.17
150000	11.92	11.93	11.95	12.08	12.25	13.63	15.46	17.41	19.46
200000	12.21	12.22	12.23	12.34	12.47	13.55	14.98	16.49	18.08

[a] These values are exact for $n_2 = t - 2n_1$ where t corresponds to the effective number of observations (in this application, the number of securities times the number of periods), and correspond to the F values reported in Table II and for the values denoted F_3 elsewhere. They are numerically accurate to one significant digit for other F values with degrees of freedom reported in this paper. For example, F_1 for the seven factor model with 1,000 and 100,000 degrees of freedom as a critical value of 10.58 as opposed to 11.97. To avoid possible ambiguity we report a critical value of 20 in Tables IV, V, and VI. The posterior odds ratio is defined for minimal prior information (see Klein and Brown [16]) and is defined in the text.

and Ross [30]. This bilinear paradigm suggests a very simple test of the arbitrage pricing model, a test that does not require that the factors common across securities be multivariate normal. We have implemented this test on the basis of the same data used by Roll and Ross, and these data are in apparent conflict with the model at standard levels of statistical significance. However, with very many observations it is possible to reject any hypothesis (except perhaps the purely artifactual) at one's favorite level of statistical significance. When we adjust the size of the test to take this into account, our results are consistent with the three factor APM. Further, we continue to reject the five and seven factor versions, a fact which actually adds support to the three factor version. In addition, the bilinear paradigm gives useful insight into the appropriate specification of the model. By examining variations in the number of factors and by organizing securities according to industrial classifications, we conclude that there are few rather than many economy wide factors that appear to be priced in an APM framework.

The fact that we are able to test a model, the APM, that imposes so very few constraints on the observed data, suggests that the bilinear paradigm may be a useful tool for the examination of more highly structured models of equilibrium asset pricing.

APPENDIX

Derivation of Results in Text

A.1 *Derivation of the Test Statistic*

Suppose we have two groups, with m_1 and m_2 securities respectively. If the equilibrium asset pricing model holds, we may write the matrix of returns on the

$m = m_1 + m_2$ securities R.

$$\underset{t \times m}{R} = \left[\underset{t \times m_1}{R_1} \middle| \underset{t \times m_2}{R_2} \right]$$

as

$$R = AB + U \tag{A.1}$$

$$\begin{aligned} [R_1 \mid R_2] &= A[B_1 \mid B_2] + [U_1 \mid U_2] \\ &= [AB_1 \mid AB_2] + [U_1 \mid U_2] \end{aligned} \tag{A.2}$$

The alternative is that the characteristics, their market prices and/or implied risk-free rates differ across groups:

$$[R_1 \mid R_2] = [A_1B_1 \mid A_2B_2] + [U_1 \mid U_2] \tag{A.3}$$

In order to test the null hypothesis implicit in the representation (A.2) against the alternative given by Equation (A.3),[32] it is possible to consider both a likelihood ratio test and a large sample F test.

To compute a likelihood ratio statistic, we need a ratio of likelihoods maximized under the alternative hypothesis (A.3) and the null hypothesis (A.1). The likelihood under the alternative hypothesis is given by[33]

$$\begin{aligned} L_{H_a} = {} & (2\pi)^{-tm_1} \mid D_1 \mid^{-t/2} \exp\left\{-\frac{1}{2}\, trD_1^{-1}[R - A_1B_1]'[R - A_1B_1]\right\} \\ & \times (2\pi)^{-tm_2} \mid D_2 \mid^{-t/2} \exp\left\{-\frac{1}{2}\, trD_2^{-1}[R - A_2B_2]'[R - A_2B_2]\right\} \end{aligned}$$

The logarithm of this likelihood can be maximized by maximizing the likelihood of each group independently. Substituting ML estimates of A, B, and D from (11) into the likelihood function above we can show after simplification that

$$\max L_{H_a} = (2\pi)^{-tm} \begin{vmatrix} \hat{D}_1 & 0 \\ 0 & \hat{D}_2 \end{vmatrix}^{-t/2} \exp\left[-\frac{1}{2}\, m\right]$$

Similarly, we find that the maximum of the likelihood under the null hypothesis is given by

$$\max L_{H_0} = (2\pi)^{-tm} \mid \hat{D} \mid^{-t/2} \exp\left[-\frac{1}{2}\, m\right]$$

[32] Note that, even if, as in the APM context, individual elements of A_1 and A_2 are not identified, it is not possible to find nonsingular matrices G_1 and G_2 such that for arbitrary $A_1 \neq A_2$, $A_1G_1 = A_2G_2 = A$.

[33] Here and in the subsequent development, we shall for expositional purposes assume that the covariance matrix of errors is diagonal. If, as in the CAPM context, there were a full covariance matrix $\sum$, the analysis follows if we post-multiply the error matrix $R - AB$ by the matrix Γ where $\Gamma\Gamma' = \sum^{-1}$. The matrix Γ has a simple form if, for example, the errors themselves have a factor structure. Note that for $\sum$ diagonal, Γ is diagonal. It is true that the diagonal covariance matrix facilitates estimation but this assumption is not otherwise required in the definition of the test statistics.

where we take both groups together. So the ratio of maximized likelihoods is

$$\lambda = \frac{|\hat{D}_u|^{-t/2}}{|\hat{D}_c|^{-t/2}} = |\hat{D}_c\hat{D}_u^{-1}|^{t/2} \tag{A.4}$$

where $\hat{D}_c$ and $\hat{D}_u$ represent the ML estimate of constrained and unconstrained diagonal error covariance matrices. It is known (e.g., Silvey [35] p. 113) that in large samples the likelihood ratio statistic

$$-2 \ln \lambda = t \ln|\hat{D}_c\hat{D}_u^{-1}| \tag{A.5}$$

is distributed as Chi-square with degrees of freedom the number of restrictions required to go from the alternative to the restricted hypothesis; in this case, the degrees of freedom is given by r, the product of the number of rows and number of columns of A.[34] Where the ratio of constrained to unconstrained variances is close to unity note that

$$-2 \ln \lambda = t \ln|\hat{D}_c\hat{D}_u^{-1}| \simeq t \cdot tr\hat{D}_c\hat{D}_u^{-1} - mt \tag{A.6}$$

There is a simple large sample F statistic interpretation of this test. Assume that the matrices D and B are known, and that the matrices A and (A_1, A_2) are estimated by constrained and unconstrained least squares respectively, accounting for the heteroskedasticity correction D.

The residuals from the t cross-section GLS regressions will be given by

$$\hat{U}_c = R - \hat{A}B$$

$$\text{where} \qquad \hat{A} = RD^{-1}B'[BD^{-1}B']^{-1}$$

for D the diagonal cross-section covariance matrix of errors. The sum of squared errors after this heteroskedasticity correction will be given by

$$SS_{c\tau} = \hat{\mathbf{u}}_{c\tau}D^{-1}\hat{\mathbf{u}}'_{c\tau}$$

where $\hat{\mathbf{u}}_{c\tau}$ is the τ^{th} row of $\hat{U}_c$. Summing over periods we have

$$SS_c = \textstyle\sum_{\tau=1}^{t} \hat{\mathbf{u}}'_{c\tau}D^{-1}\hat{\mathbf{u}}'_{c\tau}$$

Using the properties of the trace operator (e.g. Theil [38] p. 15) we write this sum as

$$\begin{aligned} SS_c &= trD^{-1}\hat{S}_c \quad \text{for} \quad \hat{S}_c = \hat{U}_c'\hat{U}_c \\ &= t \cdot tr\hat{D}_cD^{-1} \end{aligned}$$

where $\hat{D}_c$ is a diagonal covariance matrix comprising elements estimated from the diagonal of $\hat{S}_c/t$.

In a similar fashion, the unconstrained sum of squares will be given by

$$\begin{aligned} SS_u &= \textstyle\sum_{\tau=1}^{t} \hat{\mathbf{u}}_{u\tau.}\, D^{-1}\hat{\mathbf{u}}'_{u\tau.} \\ &= t \cdot tr\hat{D}_uD^{-1} \end{aligned}$$

The F statistic is given by

[34] Note that the null hypothesis constrains $A_1 = A_2 = A$ whereas the alternative allows the elements of A_2 to be free parameters.

$$F_{r,(tm-2r)} = \frac{(SS_c - SS_u)/r}{SS_u/(tm - 2r)}$$

$$= \frac{(t \cdot tr\hat{D}_c D^{-1} - t \cdot tr\hat{D}_u D^{-1})/r}{t \cdot tr\hat{D}_u D^{-1}/(tm - 2r)} \tag{A.7}$$

If we substitute maximum likelihood estimates for B and D, using for D the estimates obtained under the alternative hypothesis (e.g., Theil [38] p. 402), the F statistic further simplifies to

$$\hat{F}_{r,(tm-2r)} = \frac{(t \cdot tr\hat{D}_c\hat{D}_u^{-1} - mt)/r}{mt/(mt - 2r)} \tag{A.8}$$

which is now only approximately distributed as F with the stated degrees of freedom.[35] However, for large t and m note that $r\hat{F}_{r,(tm-2r)}$ approximates the likelihood ratio statistic introduced earlier in Equation (A.6).[36]

How good is this approximation? In results not reported here (but available from the authors), a simulation experiment was conducted to determine the adequacy of the approximation $D = \hat{D}_c$. Two thousand returns for six securities were generated firstly from i.i.d. Normal (0,1) variables and, second, to model the effects of non-Normality from i.i.d. drawings from the empirical distribution of the value-weighted daily stock market index. Since for these examples $D = I$, it was possible to construct F values given in Equations (A.7) and (A.8). On the basis of 250 replications of each experiment, there was no measurable difference in either case between the respective F values which in both cases corresponded closely to the theoretical F distribution. This, of course, is due to the very large number of observations used in this study. Note that in small samples both the Chi-square and F tests are misspecified. Evans and Savin [4] discuss a suggested sample bias correction.

While Equation (A.8) describes exactly the F ratios reported in Table II, using values of D derived from the application of Jöreskog's algorithm (see Appendix A.2), for $\hat{D}_c$ using the sixty security results and $\hat{D}_u$ using the thirty security subgroups taken separately and $r = t \cdot k$, the other F statistics reported in the paper used a slight modification of this procedure.

For statistics F_1, F_2, and F_3, the matrix $\hat{D}_u$ was taken from the diagonal elements of the covariance matrix of estimated errors $\hat{U}_u$ from the multivariate regression used to estimate A_u:[37]

$$R = A_u\hat{B}_u + U_u \tag{A.9}$$

where

[35] Theil considers imposing constraints on the parameters of L time series regressions. Our constraints are imposed on the parameters of t cross-section regressions each of which for the unrestricted hypothesis estimates $2 \times k$ parameters from $m = m_1 + m_2$ observations.

[36] The asymptotic distributions of the likelihood ratio statistic and of $r \cdot \hat{F}_{r,(tm-2r)}$ approach one another, since F_{d_1,d_2} approaches $\chi^2(d_1)/d_1$ as d_2 approaches infinity (Johnson and Kotz [14] p. 84).

[37] $\hat{A}_u = R\hat{D}^{-1}\hat{B}'[\hat{B}\hat{D}^{-1}\hat{B}']^{-1}$ where $\hat{D}$ is estimated from the initial MLFA defining $\hat{B}$.

$$\underset{t\times 2(k+1)}{A_u} = \begin{bmatrix} \underset{t\times 1}{\lambda_{01}} & \vdots & \underset{t\times k}{A_{k1}} & \vdots & \underset{t\times 1}{\lambda_{02}} & \vdots & \underset{t\times k}{A_{k2}} \end{bmatrix}$$

$$\underset{2(k+1)\times m}{\hat{B}_u} = \begin{bmatrix} \iota' & O \\ \hat{B}_{k1} & O \\ O & \iota' \\ O & \hat{B}_{k2} \end{bmatrix} \tag{A.10}$$

and $\hat{B}_{ki}$ represent the factor loading matrices estimated by an initial Jöreskog k factor analysis. The $\hat{D}_u$ matrix thus estimated applies to a model where constant terms λ_{oi} and A_{ki} matrices differ across groups. The statistic F_1 is used to test the hypothesis that λ_{oi} is the same across groups but that the A_{ki} matrices potentially differ. The matrix $\hat{D}_{c1}$ is obtained from the sample covariance matrix of U_{c1} estimated on the basis of the multivariate regression used to estimate A_{c1}:

$$R = A_{c1}\hat{B}_{c1} + U_{c1} \tag{A.11}$$

where

$$\underset{t\times(2k+1)}{A_{c1}} = [\lambda_0 \;\vdots\; A_{k1} \;\vdots\; A_{k2}]$$

$$\underset{(2k+1)\times m}{\hat{B}_{c1}} = \begin{bmatrix} & \iota' & \\ \hat{B}_{k1} & & O \\ O & & \hat{B}_{k2} \end{bmatrix} \tag{A.12}$$

and F_1 is evaluated by

$$F_1 = \frac{(t \cdot tr\hat{D}_{c1}\hat{D}_u^{-1} - mt)/t}{mt/t(m - 2(k+1))} \tag{A.13}$$

which will be distributed approximately as $F_{t,t(m-2(k+1))}$. To construct the test of the hypothesis that the A matrix is the same across groups with λ_0 potentially differing, construct $\hat{D}_{c2}$ using

$$\underset{t\times(k+2)}{A_{c2}} = [\lambda_{01} \;\vdots\; \lambda_{02} \;\vdots\; A]$$

$$\underset{(k+2)\times m}{\hat{B}_{c2}} = \begin{bmatrix} \iota' & O \\ O & \iota' \\ & \hat{B} \end{bmatrix} \tag{A.14}$$

where $\hat{B}$ is the factor loading matrix estimated on the basis of the overall group of $m = 60$ securities. The test statistic is

$$F_2 = \frac{(t \cdot tr\hat{D}_{c2}\hat{D}_u^{-1} - mt)/tk}{mt/t(m - 2(k+1))} \tag{A.15}$$

which will be distributed approximately as $F_{tk,t(m-2(k+1))}$. Finally, to examine the hypothesis that λ_0 and a are the same across groups, D_{c3} is estimated using

$$\underset{t\times(k+1)}{A_{c3}} = [\lambda_0 \ \vdots \ A]$$

$$\underset{(k+1)\times m}{\hat{B}_{c3}} = \begin{bmatrix} \iota' \\ \hat{B} \end{bmatrix} \tag{A.16}$$

and defining

$$F_3 = \frac{(t \cdot tr\hat{D}_{c3}\hat{D}_u^{-1} - tm)/t(k+1)}{tm/t(m - 2(k+1))} \tag{A.17}$$

where F_3 is distributed approximately as $F_{t(k+1),t(m-2(k+1))}$.

A2. The Relationship between Jöreskog's Procedure and Estimation of the APM

To obtain the factor analysis model analyzed by Jöreskog from the model for the observations assumed in the APM, we require that the mean of the process generating security returns is stationary over time.

$$E_{\tau i} = \lambda_{\tau 0} + \lambda_{\tau 1}b_{1i} + \cdots + \lambda_{\tau k}b_{ki} = E_i \forall \tau, i \tag{A.18}$$

In general, this requires that the implied risk-free rate $\lambda_{\tau 0}$ is constant over time.[38] Then we may write the model for the observation as

$$\begin{aligned} R &= AB + U \\ &= \lambda_0\iota' + A_kB_k + U \\ &= \iota\mathbf{E} + \Delta_kB_k + U \end{aligned} \tag{A.19}$$

where $\mathbf{E}$ is interpreted as the vector of unconditional mean returns on the m securities, $\underset{t\times k}{\Delta_k}$ is interpreted as the unobserved realizations of the k common factors for each of t periods. $\Delta_k = [A_k - \Lambda]$ where $\Lambda = \{\lambda_{\tau j}\}$.

With this restriction the likelihood function can be written

$$L[\Delta, E, B, D \mid R] = c \cdot |D|^{-t/2}\exp\{-\tfrac{1}{2}\, trD^{-1}[R - \iota\mathbf{E} - \Delta_kB_k]'[R - \iota\mathbf{E} - \Delta_kB_k]\} \tag{A.20}$$

Just as before, it is possible to solve directly for the values B, D, and Δ which satisfy the necessary conditions for a maximum to the likelihood (A.10):

$$\begin{aligned} \hat{\mathbf{E}} &= \overline{\mathbf{R}} - \overline{\Delta}_k\hat{B}_k \\ \hat{B}_k &= [\overline{\Delta}_kM\hat{\Delta}_k]^{-1}\Delta_k'MR \\ \hat{\Delta}_k &= R\hat{D}^{-1}\hat{B}'[\hat{B}\hat{D}^{-1}\hat{B}']^{-1} \\ \hat{d}_{ii} &= s_{ii}^*/t \end{aligned} \tag{A.21}$$

where $\hat{s}_{ii}^*$ is the i^{th} diagonal element of

$$S^* = [R - \iota\hat{\mathbf{E}} - \hat{\Delta}_k\hat{B}_k]'[R - \iota\hat{\mathbf{E}} - \hat{\Delta}_k\hat{B}_k]$$

$$\overline{R} = \iota'R/t, \qquad \overline{\Delta}_k = \iota'\hat{\Delta}_k/t$$

[38] Except where $\sum b_{ji=1}$ for all i.

and M is the idempotent matrix that takes differences from means $M = [I - \iota\iota'/t]$.

The solution to the set of simultaneous equations preserves the previous interpretation of taking a set of starting values for B and D, and doing successive cross-section and time series regressions until convergence. Notice, however, that these regressions do not identify the matrix of factor prices Λ_k or the implied risk-free λ_τ.

To this point we have made no particular assumptions about the *ex ante* distribution of the elements of Δ_k. If we make the further assumption that the unobserved rows of Δ are independent drawings from a multivariate normal process, with mean zero and covariance matrix given by the identity matrix, integrating the likelihood function (10) over these unobserved elements of Δ yields the standard likelihood function used in factor analysis expressed solely in terms of E, B, and D. Maximizing this likelihood for the parameters yields the following system of equations in the maximum likelihood estimates[39]

$$\hat{\mathbf{E}} = \iota' R/t$$
$$\hat{B}\hat{D}^{-1}V = (I + \hat{B}\hat{D}^{-1}\hat{B}')\hat{B}$$
$$\hat{d}_{ii} = \hat{s}_{ii}/t \qquad i = 1, m \tag{A.22}$$

where $\hat{s}_{ii}$ is a diagonal element of $\hat{S} = V - \hat{B}'\hat{B}$, for V given by the sample covariance matrix Jöreskog [14] provides an efficient algorithm for solving this set of simultaneous equations.

To verify that the maximum likelihood solution produced by Jöreskog's procedure given as the solution to (A.22) solves the simultaneous equations that define the "direct" maximum likelihood estimates of the parameters in (A.21), substitute in (A.21) for $M\hat{\Delta}_k$, yielding after some rearrangement

$$\hat{B}\hat{D}^{-1}V\hat{D}^{-1}\hat{B}'[\hat{B}\hat{D}^{-1}\hat{B}']^{-1}\hat{B} = \hat{B}\hat{D}^{-1}V \tag{A.23}$$

noting that the sample covariance of returns is given by $V = R'MR/t$. Substituting into the left-hand side of (A.23) the expression for $\hat{B}\hat{D}^{-1}V$ from (A.22)

$$\begin{aligned}&\{[I + \hat{B}\hat{D}^{-1}\hat{B}']\hat{B}\}\hat{D}^{-1}\hat{B}'[\hat{B}\hat{D}^{-1}\hat{B}']^{-1}\hat{B}\\ &= [I + \hat{B}\hat{D}^{-1}\hat{B}']\hat{B}\\ &= \hat{B}\hat{D}^{-1}V \quad \text{from (A.22)}\end{aligned}$$

which is the right-hand side of (A.23). A similar equivalence can be found for D.

Thus, if the unconditional mean returns on the individual securities are stationary in time, then the Roll and Ross procedure yields full maximum likelihood estimates of the parameters of the bilinear model and, more importantly, these estimates do not depend on the common factors having a normal probability distribution. However, their cross-section regressions of security returns against factor loadings do not impose the stationary restriction given by Equation (A.18), and this fact alone may explain the nonstationarity by Gibbons [9].

[39] For details, see Lawley and Maxwell [20]. See also Press [27] p. 308.

REFERENCES

1. P. Brennan. "A Test of the Arbitrage Pricing Model." Unpublished memo: University of British Columbia, 1981.
2. N-F. Chen. "Some Empirical Tests of the Theory of Arbitrage Pricing." Unpublished memo: University of Chicago, 1982.
3. P. L. Cheng and R. R. Grauer. "An Alternative Test of the Capital Asset Pricing Model." *American Economic Review* 70 (1980), 660–71.
4. G. B. A. Evans and N. E. Savin. "Conflict Among the Criteria Revisited: The W, LR and LM Tests," *Econometrica* 50 (1982), 737–48.
5. E. F. Fama. *Foundations of Finance.* New York: Basic Books, 1976.
6. ——— and J. MacBeth. "Risk, Return and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (1973), 607–35.
7. A. Gehr. "Some Tests of the Arbitrage Pricing Theory." *Journal of the Mid-West Finance Association* (1973) 91–105.
8. M. R. Gibbons. "Multivariate Tests of Financial Models: A New Approach." *Journal of Financial Economics* 10 (1982), 3–27.
9. ———. "Empirical Evaluation of the Return Generating Process of the Arbitrage Pricing Theory." Unpublished memo: Stanford University, 1981.
10. A. S. Goldberger. "Unobservable Variables in Econometrics." In P. Zarembka, *Frontiers in Econometrics.* New York: Academic Press, 1974, pp. 193–213.
11. P. J. Hess. "Dividend Yields and Stock Returns: A Test for Tax Effects." Unpublished memo: Ohio State University, 1980.
12. M. C. Jensen. "Capital Markets: Theory and Evidence." *Bell Journal of Economics and Management Science* 3 (1972), 358–98.
13. J. D. Jobson. "A Multivariate Linear Regression Test for the Arbitrage Pricing Theory." *Journal of Finance* 37 (1982), 1037–42.
14. N. I. Johnson and S. Kotz. *Continuous Univariate Distributions-2.* Boston: Houghton Mifflin Company, 1970.
15. K. G. Jöreskog. "Some Contributions to Maximum Likelihood Factor Analysis." *Psychometrika* 32 (1967), 443–82.
16. R. W. Klein and S. J. Brown. "Model Selection Under Minimal Prior Information." Unpublished memo: Bell Laboratories.
17. A. Kraus and R. Litzenberger. "Skewness Preference and the Valuation of Risk Assets." *Journal of Finance* 31 (1976), 1085–1100.
18. J. B. Kruskal. "Factor Analysis and Principal Components." In W.H. Kruskal and J.M. Tanur, *International Encyclopedia of Statistics.* The Free Press, 1978, pp. 307–328.
19. L. Kryzanowski and M. C. To. "General Factor Models and the Structure of Security Returns." Unpublished memo: Concordia University, 1982.
20. D.N. Lawley and A.E. Maxwell. *Factor Analysis as a Statistical Method.* London: Butterworth and Company, 1963.
21. E. E. Leamer. *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: John Wiley & Sons, 1978.
22. D. Lindley. "A Statistical Paradox." *Biometrika* 44 (1957), 187–92.
23. J. Lintner. "Security Prices, Risk, and Maximal Gains From Diversification." *Journal of Finance* 20 (1965), 587–616.
24. M. Miller and M. Scholes. "Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings." In M. Jensen (ed.), *Studies in the Theory of Capital Markets.* New York: Praeger, 1972, pp. 47–78.
25. J. Mossin. "Equilibrium in a Capital Asset Market." *Econometrica* 34 (1966), 768–83.
26. U. Olsson. "On the Robustness of Factor Analysis Against Crude Classification of the Observations" *Multivariate Behavioral Research* 14 (1979), 485–500.
27. S. J. Press. *Applied Multivariate Analysis.* New York: Holt, Rinehart and Winston, 1972.
28. R. Roll. "A Critique of the Asset Pricing Theory's Tests. *Journal of Financial Economics* 4 (1977), 129–76.
29. ———. "A Possible Explanation of the Small Firm Effect." *Journal of Finance* 36 (1981), 879–87.

30. ——— and S. A. Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." *Journal of Finance* 35 (1980), 1073–1102.
31. B. Rosenberg and V. Marathe. "Tests of Capital Asset Pricing Hypotheses." In H. Levy, ed., *Research in Finance.* Greenwich: Jai Press, 1979, pp. 115–223.
32. S. A. Ross. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 3 (1976), 343–62.
33. J. Shanken. "Analysis of the APT and Its Empirical Testability." Unpublished memo: University of California, Berkeley, 1982.
34. W. F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19 (1964), 425–42.
35. S. D. Silvey. *Statistical Inference.* Baltimore: Penguin Books, 1970.
36. N. A. Sinclair. "Security Return Data and 'Blind' Factor Analysis" Unpublished memo: Australian Graduate School of Management, 1981.
37. R. F. Stambaugh. "On the Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis." Unpublished memo: University of Pennsylvania, 1981.
38. H. Theil. *Principles of Econometrics.* New York: John Wiley & Sons, 1971.
39. A. Zellner. *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley & Sons, 1971.