A CRITICAL REEXAMINATION OF THE EMPIRICAL EVIDENCE ON THE ARBITRAGE PRICING THEORY

By

Phoebus J. Dhrymes*
Irwin Friend**
N. Bulent Gultekin**

Working Paper No. 18-83

THE WHARTON SCHOOL

University of Pennsylvania
Philadelphia, PA 19104

* Columbia University, Department of Economics.
** The Wharton School, University of Pennsylvania.

The contents of this paper is the sole responsibility of the authors.

A CRITICAL REEXAMINATION OF THE EMPIRICAL EVIDENCE ON THE ARBITRAGE PRICING THEORY

Phoebus J. Dhrymes*
Irwin Friend**
N. Bulent Gultekin**

August 1982
Revised May 1983

Forthcoming in Journal of Finance

*Columbia University, Department of Economics
**The Wharton School, University of Pennsylvania

We owe a great deal to Mustafa N. Gultekin for his generous help in computer programming and for sharing his expertise in computer applications of factor analytic methods with us.

1. Introduction

The APT model as exposited by Ross [15], [16] and extended by Huberman [8] has provided the basis of an extensive literature, as for example, Brown and Weinstein [1], Chen [3], Hughes [9], Ingersoll [10], Jobson [11], Reinganum [12], Roll and Ross (RR) [14] and Shanken [18], to mention but a few. It has also been treated in a more general context by Chamberlain and Rothschild [12].

Our purpose here is not to comment on this collection of papers--many of which are unpublished--but to reexamine the evidence presented by RR and point out major pitfalls involved in the empirical methodology employed by them and others who have followed their lead.

RR claim that, on theoretical and more important on empirical grounds, arbitrage pricing theory (APT) is an attractive alternative to the capital asset pricing model (CAPM). The APT, it is argued, requires less stringent and presumably more plausible assumptions, is more readily testable since it does not require the measurement of the market portfolio, and may be better able to explain the anomalies found in the application of the CAPM to asset returns.

The acceptability of the APT, like that of the CAPM or any other theory, ultimately depends on its ability to explain the relevant empirical evidence. Our subsequent discussion will show that many of the problems associated with the CAPM are also present in the APT context and that its ability to explain the relevant empirical evidence is not markedly superior. Moreover, it raises questions not only about the RR and related empirical investigations of the APT but also about the testability of that theory in the present state of the art. A major part of the problem results from the necessity to break down the "universe" being analyzed through the APT model; this is forced upon the investigator by the fact that computer software does not permit factor analysis involving covariance or correlation matrices of high order (in the case of RR 1260).

Our paper has three major thrusts:

1. It is generally impermissible to carry out tests on whether a given "risk factor is priced"; this is a result that is inherent in the structure of standard factor analytic models, unless one is prepared a priori to specify that in the "true" structure certain factor loadings are known.
2. Factor analyzing small groups of (30) securities is not equivalent to factor analyzing a group of securities sufficiently large for the APT model to hold.
3. As one increases the size of the security groups to which the APT/factor analytic procedures are applied, the number of "factors" determined increases.

2. The APT Model: Implications and the Nature of Empirical Tests

In order to establish notation and make this paper as self-contained as possible, we give a brief exposition of the APT model and discuss some aspects of its empirical implications.

The model begins by postulating the return generating function

$$r_{t} = E_{t\cdot} + f_{t\cdot}B + u_{t\cdot} \tag{1}$$

where $r_{t\cdot}$ is an m-element row vector containing the observed rates of return at time t for the m-securities under consideration; $E_{t\cdot}$ is similarly an m-element row vector containing the expected (mean) returns at time t. Finally,

$$v_{t\cdot} = f_{t\cdot}B + u_{t\cdot} \tag{2}$$

represents the error process at time t. It is an essential feature of the APT model that the error process has two components: the idiosyncratic component

$$u_t, \quad t = 1, 2, \ldots$$

and the common component

$$f_{t\cdot}B .$$

It is assumed that

$$\{u'_{t\cdot}: \ t = 1, 2, \ldots\}$$

is a sequence of independent identically distributed (i.i.d.) random vectors with

$$E(u'_{t\cdot}) = 0 , \ Cov(u'_{t\cdot}, f'_{t\cdot}) = 0 \ \text{ and } \ Cov(u'_{t\cdot}) = \Omega , \tag{3}$$

the covariance matrix Ω being <u>diagonal</u> and such that

$$0 < \omega_{ii} < \infty , \quad i = 1, 2, \ldots m . \tag{4}$$

Regarding the common component, we note that the form in which it is stated creates an identification problem, since neither $f_{t\cdot}$ nor B are directly observable. We (partly) eliminate this problem by specifying that

$$\{f'_{t\cdot}: \ t = 1, 2, \ldots\}$$

is a sequence of k-element i.i.d. random vectors with[1]

$$E(f'_{t\cdot}) = 0, \ Cov(f'_{t\cdot}) = I . \tag{5}$$

It is a consequence of the assertions above that

$$\left\{(r_{t\cdot} - E_{t\cdot})': \ t = 1, 2, \ldots\right\}$$

is a sequence of i.i.d. random vectors with

$$E\left[(r_{t\cdot} - E_{t\cdot})'\right] = 0, \ Cov\left[(r_{t\cdot} - E_{t\cdot})'\right] = B'B + \Omega = \Psi . \tag{6}$$

We further note that

$$\begin{aligned} \mathrm{Cov}(r_{ti}, r_{tj}) &= b'_{\cdot i} b_{\cdot j} & i \neq j \\ &= b'_{\cdot i} b_{\cdot i} + \omega_{ii} & i = j \end{aligned} \quad (7)$$

and that indeed the columns of B ($b_{\cdot i}$, which are k×1) contain information on the covariation of securities.

Finally, since only Ψ can be estimated directly from the data, (6) shows that there is a further identification problem not eliminated by the assertion in (5) and the discussion of footnote 1. For if B is a matrix satisfying (6) and (1) and even if $f_{t\cdot}$ obeys (5), then for any orthogonal matrix Q, QB also satisfies (6) and (1). Hence, B can be identified only up to left multiplication by an orthogonal matrix, unless we are prepared to put a priori restrictions on the elements of B.

Now, just what restrictions on empirical evidence are implied by the APT model? First, we should note that the proof of the crucial implication of APT requires the invocation of a strong law of large numbers (SLLN); hence, the universe of securities to which one seeks to apply the model must contain a sufficiently large number of them so that the invocation of the SLLN may be reasonably justified.

Secondly, the fundamental conclusion of APT requires that there exist a (k + 1)-element row vector, $c_{t\cdot}$, such that

$$E_{t\cdot} = c_{t\cdot} B^*, \qquad t = 1, 2, \ldots, T \quad (8)$$

where

$$B^* = \begin{pmatrix} e' \\ B \end{pmatrix}, \quad (9)$$

e being an m-element column of ones.

Thus, the no arbitrage condition characterizing equilibrium rates of return requires that if $x_{t\cdot}$ is such that

$$u_{t\cdot}x'_{t\cdot}$$

is an entity to which the SLLN applies and if $x_{t\cdot}$ belongs to the column null space of B^*, then $E_{t\cdot}$ <u>must lie in the row space of B^*</u>.[2] If the number of securities (m) is sufficiently large, then for any desired degree of approximation we can rewrite (1) as

$$r_{t\cdot} = c_{t\cdot}B^* + f_{t\cdot}B + u_{t\cdot}, \qquad t = 1, 2, \ldots, T . \tag{10}$$

The restriction on empirical evidence imposed by (10) is rather stringent; in particular, it requires that no other (relevant) economic/financial variables have any bearing on the determination of expected rates of return, once the impact (on the expected rates of return) of the covariation among securities is taken into account.

3. Significance Tests for the "Priced" Risk Factors

The empirical tests of APT carried out by RR and others are based on a two step factor analytic approach. Factor analytic methods are, in effect, suggested by the formulation in (1) and the composition of the covariance matrix in (6). In the first step one determines the number of factors (k) and estimates the elements of B and in the second stage, using the latter as the "independent variables," one estimates the vector $c_{t\cdot}$, whose elements have the interpretation that c_{ti} is the risk premium attached to the i^{th} factor, $i = 1, 2, \ldots, k$, while c_{to} is the risk-free rate, or possibly the return on a zero-beta asset.

The question often arises as to whether all (common) risk factors are priced or only a subset thereof. In this section, we wish to address the methodological issues bound up with these concerns.

Thus, suppose T is sufficiently large so that these covariance or correlation matrices, Ψ, can be estimated with reasonable accuracy and by factor analysis we estimate B, say by $\tilde{B}$, and Ω, say by $\tilde{\Omega}$. Thus, we have implicitly estimated

$$\tilde{\Psi} = \tilde{B}'\tilde{B} + \tilde{\Omega} \quad , \tag{11}$$

which completes the first stage; in the second stage, for each t, we may estimate

$$\tilde{c}'_{t\cdot} = (\tilde{B}^*\tilde{\Psi}^{-1}\tilde{B}^{*\prime})^{-1}\tilde{B}^*\tilde{\Psi}^{-1}r'_{t\cdot}, \qquad t = 1, 2, \ldots, T \; . \tag{12}$$

If the underlying error process is normal and if T is sufficiently large, it is easy to show that (approximately)

$$(\tilde{c}_{t\cdot} - c_{t\cdot})' \sim N\left[0, \; (B^*\Psi^{-1}B^{*\prime})^{-1}\right] \; . \tag{13}$$

Thus, we may view the $\{\tilde{c}'_{t\cdot}: \; t = 1, 2, \ldots, T\}$, approximately, as drawings from a multivariate normal distribution with mean[3]

$$c'_{t\cdot}: \quad t = 1, 2, \ldots, T$$

and covariance matrix

$$(B^*\Psi^{-1}B^{*\prime})^{-1} \; .$$

Recalling, however, that B is identified by factor analytic procedures only to the extent of left multiplication by an orthogonal matrix, we are led to doubt the manner in which tests on individual elements of $c_{t\cdot}$ are typically carried out. General tests, see for example RR, involve the introduction of other explanatory variables and a test of the hypothesis that the corresponding coefficients are zero.

In the remainder of this section, we shall examine the question whether, in the context of the specification

$$r_{t\cdot} = c_{t\cdot}B^* + d_{t\cdot}P + v_{t\cdot} \tag{14}$$

where P is a matrix of "extraneous" variables and B^* is only identified to within left multiplication by

$$Q^* = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \tag{15}$$

and Q is orthogonal, it is possible to have unambiguous tests of significance on $d_{t\cdot}$, c_{to} and c_{ti}, $i = 1, 2, \ldots, k$. To avoid extraneous issues, we shall suppose that T is sufficiently large so that the estimates of B obtained by a factor analytic approach have negligible sampling variation and we can deal with them as if they were their probability limits. Let B_o be the "true" parameter matrix as initially specified in (1). Then the output of the factor analytic procedure and therefore the set of explanatory variables in (14) is a matrix, say, $\bar{H}$, which is related to the true matrix H_o through

$$\bar{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & I \end{bmatrix} H_o \quad , \quad H_o = \begin{bmatrix} e' \\ B_o \\ P \end{bmatrix} \tag{16}$$

Note that H_o is unambiguously specified although B_o is unknowable; what is knowable is only the transformation

$$\bar{B} = QB_o$$

where Q is an arbitrary orthogonal matrix. Hence, the only unambiguous conclusions that may be derived from such an analysis must be conclusions "modulo" Q, i.e., conclusions that do not in any way depend on Q.

We have

Proposition 1: Consider the general model in (14) and suppose T is sufficiently large so that sampling variation may be ignored. Define

$$h_{t\cdot} = (H_o\Psi^{-1}H_o')^{-1}H_o\Psi^{-1}r_{t\cdot}' \ , \ h_{t\cdot} = (c_{t\cdot}, d_{t\cdot}) \ , \ \bar{h}_{t\cdot}' = \bar{Q}h_{t\cdot}' \ , \ \bar{H} = \bar{Q}H_o$$

$$\tilde{\bar{h}}_{t\cdot} = (\bar{H}\Psi^{-1}\bar{H}')^{-1}\bar{H}\Psi^{-1}r_{t\cdot}' \ , \quad \bar{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & I \end{bmatrix} \tag{17}$$

where Q is an orthogonal matrix of order k, I is the identity matrix of order s (s being the number of extraneous explanatory variables - the rows of P), and H_o as in Equation (16); then

$$\tilde{\bar{h}}_{t\cdot}' = \bar{Q}\tilde{h}_{t\cdot}' \tag{18}$$

Proof: Obvious by noting that

$$\bar{Q}'\bar{Q} = I \ ,$$

i.e., $\bar{Q}$ is also an orthogonal matrix (of order $s + k + 1$).

It is evident that the "risk free" or "zero-beta" rate is uniquely estimated since in the obvious notation

$$\tilde{\bar{c}}_{to} = \tilde{c}_{to} \ .$$

Moreover, the vector $d_{t\cdot}$ is also uniquely estimated since

$$\bar{\tilde{d}}_{t\cdot} = \tilde{d}_{t\cdot} \ .$$

On the other hand

$$(\bar{\tilde{c}}_{t1}, \bar{\tilde{c}}_{t2}, \ . \ . \ ., \bar{\tilde{c}}_{tk})' = Q(\tilde{c}_{t1}, \tilde{c}_{t2}, \ldots, \tilde{c}_{tk})'$$

which states that, in general,

$$\bar{\tilde{c}}_{ti} \neq \tilde{c}_{ti}, \qquad i = 1, 2, \ . \ . \ ., k \ .$$

Hence, the interpretation of the regression estimates $\bar{\tilde{c}}_{ti}$ as the "risk premium" for the i^{th} common risk factor is a serious overreaching of the empirical evidence.

In RR as well as Hughes [9] among others, many tests are carried out as to how many factors are being "priced"; in RR tests are also carried out on the "significance" of individual extraneous variables as a "test" of the APT model. To what extent are the results of such tests unambiguous? This is, in part, answered by

<u>Proposition 2</u>: Under the conditions of Proposition 1 the covariance matrix of the estimator $\bar{\tilde{h}}'_{t\cdot}$ has the following properties:

1. the variance of $\bar{\tilde{c}}_{to}$ is exactly the variance of $\tilde{c}_{to}$, i.e., it does not depend on the matrix Q;
2. the covariance matrix corresponding to $\bar{\tilde{d}}_{t\cdot}$ is exactly the covariance matrix of $\tilde{d}_{t\cdot}$, i.e., it does not depend on the matrix Q;
3. the covariance matrix corresponding to the "risk premia" assigned to the common risk factors does depend on the matrix Q.

Proof: The covariance matrix of $\bar{\tilde{h}}'_{t\cdot}$ is evidently given by

$$(\bar{H}\Psi^{-1}\bar{H}')^{-1} = \bar{Q}(H_o\Psi^{-1}H_o')^{-1}\bar{Q}' \quad .$$

Thus the variance of $\bar{c}_{to}$ is simply the (1, 1) element of $(H_o\Psi^{-1}H_o')^{-1}$ which is independent of Q. Specifically, it is

$$[e'\Psi^{-1}e - e'\Psi^{-1}H_2^{o\prime}(H_2^o\Psi^{-1}H_2^{o\prime})^{-1}H_2^o\Psi^{-1}e]^{-1}$$

where

$$H_2^o = \begin{vmatrix} B_o \\ P \end{vmatrix} .$$

For the second statement of the proposition we note that the corresponding covariance matrix is

$$[P\Psi^{-1}P' - P\Psi^{-1}B_o^{*\prime}(B_o^*\Psi^{-1}B_o^{*\prime})^{-1}B_o^*\Psi^{-1}P']^{-1}$$

which evidently does not depend on Q; B_o^* is the true parameter matrix, B_o, augmented by a column of ones as in (9). For the last part of the statement we note that the relevant covariance matrix is

$$\Phi_{22} = Q[B_o\Psi^{-1}B_o' - B_o\Psi^{-1}P^{*\prime}(P^*\Psi^{-1}P^{*\prime})^{-1}P^*\Psi^{-1}B_o']^{-1}Q' \qquad (19)$$

or

$$\Phi_{22} = Q\Phi_{22}^oQ'$$

where

$$P^* = \begin{vmatrix} e' \\ P \end{vmatrix}$$

and Φ_{22}^o has the obvious meaning. This makes the dependence clear and completes the proof of the Proposition.

The results above should make it evident that tests on individual coefficients, c_{ti}, or subsets of risk premia coefficients are not unambiguous. To demonstrate this let us assume normality of the error terms so that the test statistics become unambiguously determined in their distribution. Such linear tests involve consideration of quantities like

$$A\bar{c}^{*o\prime}_{t\cdot}$$

where

$$\bar{c}^{*o}_{t\cdot} = (\bar{c}^{o}_{t1}, \bar{c}^{o}_{t2}, \ldots, \bar{c}^{o}_{tk})$$

A is a suitable matrix with known elements and $\bar{c}^{*o}_{t\cdot} = c^{*o}_{t\cdot}Q'$, $c^{*o}_{t\cdot}$ being the "true" risk premia vector. The usual test statistic then would obey, for the test of $A\bar{c}^{*o\prime}_{t\cdot} = 0$,

$$\tilde{\bar{c}}^{*}_{t\cdot}A'(A\Phi_{22}A')^{-1}A\bar{c}^{*\prime}_{t\cdot} = \tilde{c}^{*}_{t\cdot}Q'A'(AQ\Phi^{o}_{22}Q'A')^{-1}AQ\tilde{c}^{*\prime}_{t\cdot} \sim \chi^2_r$$

where

$$\Phi_{22} = Q'\Phi^{o}_{22}Q \quad \text{and} \quad A \text{ is } r\times k \; (r \leq k) \text{ of rank } r.$$

Since in general

$$Q'A'(AQ\Phi^{o}_{22}Q'A')^{-1}AQ \neq A'\left(A\Phi^{o}_{22}A'\right)^{-1}A \;,$$

a test of the hypothesis

$$A\bar{c}^{*o\prime}_{t\cdot} = 0$$

is not equivalent to a test of

$$Ac^{*o\prime}_{t\cdot} = 0$$

In the special case, however, when A is the identity matrix of order k (i.e., when we simultaneously test all risk premia), we find that the test statistic reduces to

$$\tilde{c}^*_{t\cdot}(\Phi^o_{22})^{-1}\tilde{c}^{*\prime}_{t\cdot} \sim \chi^2_k \tag{20}$$

which is appropriate for testing the hypothesis

$$c^{*o'}_{t\cdot} = 0 .$$

Thus, in this context the crucial testable hypothesis is how many factors there are and whether none of them is priced, rather than whether some of them are priced and others are not.

4. A Critical Appraisal of the RR and Similar Empirical Tests

The APT model as explained earlier requires the set of securities to which the (rates of) return generating function in (1) applies to be large enough so that a SLLN applies. It has the important implication that

$$E_{t\cdot} = c_{t\cdot}B^* .$$

There is no presumption of time stationarity with respect to the vector $c_{t\cdot}$ which describes the dependence of the mean (expected) rate of return vector on the rows of B^*.

It is a practical necessity, and was explicitly assumed earlier, that the distributions of $f_{t\cdot}$ and $u_{t\cdot}$ be time stationary, i.e., it was explicitly assumed that their distribution (or at least their second moments) did not vary with t. At least, this must be so over a sufficiently long period to permit the estimation of the relevant covariance matrix.[4]

A second important implication of the APT model as exposited above is that, within any degree of (probabilistic) approximation desired, the vector of returns of the m securities in question may be written as

$$r_{t\cdot} = c_{t\cdot}B^* + f_{t\cdot}B + u_{t\cdot}, \qquad t = 1, 2, \ldots, T \tag{21}$$

which means that if we are to subject the model to empirical testing, we ought to treat <u>all m securities symmetrically</u>.

It has been a practice initiated in the paper by RR, and by now quite widely adopted by other researchers, to divide the universe of securities into a number of groups (42 in the case of RR) and treat them as "cross sections" from a population in the manner one treats a sample of households in the context of a consumer expenditures survey. The analogy is, of course, quite appealing, which is the reason for its wide acceptance. It is, however, very misleading. What enables us to use the cross-sectional information of a consumer expenditures survey to infer something about the parameters that characterize that particular universe is that each individual in the cross-section has some fairly well-defined attributes which can be measured unambiguously and independently of how many individuals there are in the cross-section coupled with a presumption of parametric homogeneity among the entities of the relevant universe.

A reflection on the nature of the model as exhibited in (21) will disclose that if we partitioned the universe into 42 groups as RR do, then for the i^{th} group we should have

$$r_{t\cdot}^{(i)} = c_{t\cdot}^{(i)} B^*_{(i)} + f_{t\cdot} B_{(i)} + u_{t\cdot}^{(i)} , \qquad i = 1, 2, \ldots, 42 \qquad (22)$$

where $r_{t\cdot}^{(1)}$ consists of the first 30 elements of $r_{t\cdot}$, $r_{t\cdot}^{(2)}$ of the second thirty elements and so on. Similarly, $B_{(1)}$ consists of the first 30 columns of B, $B_{(2)}$ of the second 30 columns and so on. The same is true of $B^*_{(1)}$, $u_{t\cdot}^{(1)}$, $B^*_{(2)}$, $u_{t\cdot}^{(2)}$, etc.

The question arises whether each group can be dealt with in isolation with any degree of assurance regarding the reliability of the ensuing results. In their paper, RR and those who have followed their lead such as, for example Hughes [9], make only the caveat that while only a relatively small number of factors may be identified in each group, one must bear in mind that perhaps it may be different factors that correspond to different groups. The remainder of their discussion appears to ignore this point and proceeds as if the same small number of factors is identified for each group; RR are particularly pleased that the number of significant factors extracted in each of their 42 groups (of alphabetically arranged 30 securities) ranges between three and five.

Unfortunately, the situation is far more grave than the literature has thus far allowed for. Treating each group of 30 securities as a cross-section and looking to the results from such an exercise for confirmatory evidence about the number of factors is not appropriate. Most importantly it should be stressed that, in general, what is the equivalent of the explanatory variables (attributes) for the 30 securities cannot be measured reliably _independently_ of the issue of how many securities we treat simultaneously. This is so since those "explanatory variables" are given by the (sub) matrix $B^*_{(i)}$ and this cannot be measured reliably in a 30-securities context--for reasons we shall explain below. Contrast this to the consumer expenditures survey context in which each individual household's income, size, composition and other relevant socioeconomic attributes can be accurately ascertained independently of how many households there are in the sample (cross-section).

The fundamental reason for this state of affairs is, essentially, that there is no necessary and simple connection between "factors" found in a 30-security context and "factors" to be determined in the context in which APT models may be held to apply. As pointed out in another context by one of the authors (Dhrymes [5]), the "factors" postulated in such models are not _ab initio_ tangible concrete entities. Rather, the structure simply represents the researchers' rendition of the well-established pattern of covariation among securities' returns. If, after prolonged empirical investigations the number of "factors" found is stabilized and an economic/financial interpretation is attached to them, we may, at that stage, think of such risk factors as reflecting fundamental economic forces at work in the securities markets. Thus, the nature of the factors is to be inferred from the observed behavior of security returns -- not the other way around. Consequently, since the "theory" indicates that this pervasive pattern of covariation affects certain groups of securities obeying certain requirements, their existence and identification can only be determined in that context.

A simple example from the psychometrics literature may help to put this matter in more intuitive terms. Factor analysis has been often applied to the results of different intelligence tests in order to determine the "common factors" underlying the components of human intelligence. The number of securities in the context of the APT is like the number of tests one administers; the number of trading days is like the number of individuals to whom one administers the tests. If there are too few of them (individuals), the confidence one has in the results is not very strong; if there are very many, then one's confidence in the result of the analysis increases. The number of tests has nothing to do with sampling variation; it has something to do with one's universe of discourse. If a specific number of tests elicit a performance governed by the operation of k factors, then one has to analyze all of these test results (on all individuals) simultaneously. Recall that factors are not observed or observable directly; they are inferred only by the performance of individuals on all the tests administered. If only some tests are analyzed (such as using only the tests measuring verbal skills while excluding math and other relevant tests), then part of the information on how these factors affect the results is missing and hence the results of the analysis may be misleading. Using an abbreviated environment would necessarily distort the evidence, much in the manner of the fable of the blind man feeling a large elephant.

By way of a logical explanation as to why $B^*_{(i)}$ cannot be measured reliably in the context employed by RR and others, such as Brown and Weinstein [1] for example, consider the model in (10) and suppose $r_{t\cdot}$ is stated in terms of standard deviates, i.e., we subtract from each r_{ti} its mean and divide by its standard deviation, thus conforming to the standard procedures in factor analysis computer software. Making allowance for this construction, the interpretation of Ψ in (6) is now that of a correlation matrix so that its diagonal elements are unity.

Partitioning Ψ in accordance with the RR scheme above we have

$$\Psi = [\Psi_{ij}], \qquad i, j = 1, 2, \ldots, 42 \tag{23}$$

so that Ψ_{ii} is the correlation matrix for the i^{th} group of 30 securities and Ψ_{ij}, $i \neq j$, the "cross correlation" matrix between securities in the i^{th} and j^{th} groups. If we subject all 1260 securities to factor analysis simultaneously, we shall obtain estimates of B and Ω, say $\tilde{B}$ and $\tilde{\Omega}$, obeying

$$\begin{aligned} \operatorname{diag}(S) &= \operatorname{diag}(\tilde{B}'\tilde{B} + \tilde{\Omega}) \\ [\tilde{\Omega}^{-1/2}(S - \tilde{\Omega})\tilde{\Omega}^{-1/2}]\tilde{\Omega}^{-1/2}\tilde{B}' &= \tilde{\Omega}^{-1/2}\tilde{B}'(\widetilde{B\Omega^{-1}}\tilde{B}') \ . \end{aligned} \tag{24}$$

where S is the sample correlation matrix. On the other hand, factor analyzing each of the 42 groups we obtain

$$\begin{aligned} \operatorname{diag}(S_{ii}) &= \operatorname{diag}(\hat{B}_i'\hat{B}_i + \hat{\Omega}_i) \\ [\hat{\Omega}_i^{-1/2}(S_{ii} - \hat{\Omega}_i)\hat{\Omega}_i^{-1/2}]\hat{\Omega}_i^{-1/2}\hat{B}_i' &= \hat{\Omega}_i^{-1/2}\hat{B}_i'(\hat{B}_i\hat{\Omega}_i^{-1}\hat{B}_i') \end{aligned} \tag{25}$$

for i = 1, 2, . . ., 42.

In (24) and (25) we impose, respectively, the conditions that $(\tilde{B}\tilde{\Omega}^{-1}\tilde{B}')$ and $(\hat{B}_i\hat{\Omega}_i^{-1}\hat{B}_i')$, i = 1, 2, . . ., 42 be diagonal. The procedure in (25) is essentially the RR procedure, and it may be rationalized by assuming that there exist orthogonal matrices, Q_i, such that

$$\tilde{B}_i' = Q_i\hat{B}_i' \quad , \qquad i = 1, 2, \ldots, 42 \tag{26}$$

and that while the $\hat{B}_i$, $\hat{\Omega}_i$ of (25) do not satisfy (24), there is still a well defined relationship between them as given by (26). If that is, indeed, the case, then certain aspects of the RR methodology will not be inappropriate, even if the procedure would not be the most efficient possible. Intuitively, this is not very likely to be true since we seem to be arguing that the characteristic vectors corresponding to the k largest characteristic roots of a matrix and its principal submatrices are related in the manner of (26). We also seem to be arguing that ignoring the off diagonal blocks constitutes a misspecification that entails no cost.

Since the impression that (26) is correct is rather widespread, it is appropriate for us to address this issue.[5] A particularly simple way of doing this is by noting that if (26) is correct, then the length (inner product) of the columns of $\tilde{B}_i'$ and $\hat{B}_i'$ must be identical. This identity is not a matter of statistical significance and departures from it (beyond round-off errors) cannot be attributed to sampling error. Rather, the equality of the length (inner product) of the two sets of vectors is a simple consequence of the mathematical relationship noted in Equation (26); thus, any discrepancy between the lengths (inner product) of these two sets of vectors can only mean that the relationship exhibited in (26) is false. Consequently, there is no such simple relationship linking the "factor loadings" obtained from a set of, say, 8 groups each containing 30 securities and the factor loadings obtained when we "factor analyze" these 240 securities as a simple group. In order that we may explore these issues, we extract "five factors" from a group of 240 securities and "five factors" from each of its 8 constituent groups of 30 securities each. We present in Table 1 the length of the columns of $\tilde{B}_i'$ and $\hat{B}_i'$ (five in each case). The first column refers to the case where the $\hat{B}_i'$ have been obtained by the factor analysis of the 8 groups of 30 securities each; the second column refers to the case where the $\tilde{B}_i$ are the 5x30 submatrices of the matrix $\tilde{B}$ obtained by factor analyzing the entire group of 240 securities. Although this is not a particularly stringent test, the results show quite unambiguously that no such relation as (26) exists in the current context. It is quite important to realize that factor analyzing 30 security groups <u>is not equivalent in any way to factor analyzing a 240 or a 1260 security group if we impose the condition of extracting the same number of factors in the two contexts.</u> Thus, the RR procedure is not equivalent to testing the universe of securities for which the APT is postulated to hold. The number of securities used in testing the APT should be sufficiently large so that the SLLN can be invoked as shown by Ross [15] and Huberman [8].

5. A Partial Test for the Loss of Information Entailed by the RR Procedures

As indicated in the preceding section, the RR methodology does not use the information contained in the off-diagonal blocks of the matrix (23) (i.e. the covariance matrix of the *entire* set of securities). This methodology can then be justified only as a special case when the information in the off-diagonal blocks of (23) can be ignored. One may therefore interpret the RR methodology as operating on the implicit assumption that the off diagonal blocks of the matrix in (23) obey

$$\Psi_{ij} = 0, \quad i \neq j, \quad i, j = 1, 2, \ldots, 42 \tag{27}$$

and extracting five factors from each group (or 210 in all) rather than five from the entire group.

While we cannot actually test the null hypothesis (16) that $\Psi_{ij} = 0$ for the entire set of securities, we can carry out a "test" for a set of 240 securities. The (log) likelihood function (LF) under H_o as in (27) is given by

$$L_1 = -\frac{30nT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^{n} \left[\ln|B_i'B_i + \Omega_i| + \mathrm{tr}\Psi_{ii}^{-1}S_{ii}\right] \tag{28}$$

where

$$\Psi_{ii} = B_i'B_i + \Omega_i, \qquad S_{ii} = \frac{1}{T}\sum_{t=1}^{T} r_{t\cdot}^{(i)'} r_{t\cdot}^{(i)}, \qquad i = 1, 2, \ldots, n \tag{29}$$

(and as required in most computer software the return on the j^{th} security, r_{tj}, is stated as a standardized deviate), where n is the number of groups of securities (8 in our case).

Maximizing under H_o, and extracting five "factors" per group yields the RR results. In the particular case under consideration we find

$$\max_{H_o} L_1 = -\frac{30nT}{2}\left[\ln(2\pi) + 1\right] - \frac{T}{2}\sum_{i=1}^{n} \ln\left|\hat{B}_i'\hat{B}_i + \hat{\Omega}_i\right| \tag{30}$$

Under the alternative, the LF is

$$L_2 = -\frac{30nT}{2}\ln(2\pi) - \frac{T}{2}\ln\left|B'B + \Omega\right| - \frac{T}{2}\mathrm{tr}\Psi^{-1}S$$

where now

$$S = \frac{1}{T}\sum_{t=1}^{T} r_{t\cdot}' r_{t\cdot}$$

Maximizing under the alternative and still extracting five "factors" yields the RR results as would be obtained for the entire set of 240 securities. In particular, we find

$$\max_{H_1} L_2 = -\frac{30nT}{2}\left[\ln(2\pi) + 1\right] - \frac{T}{2}\ln\left|\tilde{B}'\tilde{B} + \tilde{\Omega}\right| \tag{31}$$

Now had we extracted only five "factors" from all 240 securities in the likelihood function of (28), as a nested set of hypotheses would require, we would find a maximized value for the LF in (28), say L*, obeying

$$L_1^* \leq \max_{H_0} L_1$$

Thus, if we treat

$$\ln\lambda = \max_{H_0} L_1 - \max_{H_1} L_2 = -\frac{T}{2}\left[\sum_{i=1}^{n} \ln\left|\hat{B}_i'\hat{B}_i + \hat{\Omega}_i\right| - \ln\left|\tilde{B}'\tilde{B} + \tilde{\Omega}\right|\right]$$

as the likelihood ratio test statistic, we would obtain a value for λ which is larger than the correct one. Hence

$$-2\ln\lambda = T\left[\sum_{i=1}^{n} \ln\left|\hat{B}_i'\hat{B}_i + \hat{\Omega}_i\right| - \ln\left|\tilde{B}'\tilde{B} + \tilde{\Omega}\right|\right] \tag{32}$$

would yield a value which is smaller than the correct one. Hence, if we reject on the basis of (32), we would certainly reject on the basis of the correct LR test statistic. Now, in the present case

$$n = 8, \qquad T = 2196$$

$$\sum_{i=1}^{8} \ln\left|\hat{B}_i'\hat{B}_i + \hat{\Omega}_i\right| = -15.49002$$

$$\ln\left|\tilde{B}'\tilde{B} + \tilde{\Omega}\right| = -24.16200$$

Hence

$$-2\ln\lambda = 20{,}049 \tag{33}$$

and we remind the reader that, for the correct test, the statistic would be distributed as chi-square with 1050 degrees of freedom, and the number appearing on the right side of (33) would be larger. We could, of course, continue with the formalities and actually carry out the test; but the point is abundantly made by the preceding, viz. that there is an enormous loss of information by dealing, *seriatim*, with groups of 30 securities and consequently that the RR procedure cannot be rationalized in terms of ignoring information of little value. Indeed, if we were to accept the proposition that

$$\Psi_{ij} = 0 \qquad i \neq j ,$$

then since the assignment of securities to groups is arbitrary we would conclude that there are no common risk factors thus denying the *raison d'être* of the APT model.

6. How Many Factors Are There?

Notwithstanding the criticism of the basic RR approach developed in the earlier section, we deem it important to reexamine the results obtained by them with a view to reassessing the evidence presented thus far on behalf of the empirical validity of the APT model. This is done for two reasons. First, it is an important scientific axiom that potentially important new empirical results be subjected to the test of replication; this is particularly important in the case of the work by RR since while their findings are very provocative, their data exhibit a rather vexing missing observations problem. Second, a reexamination of part of the evidence presented by them would certainly permit us to examine, in an empirical context, some of the issues we had raised earlier, and either confirm or raise doubt about certain purported empirical regularities obtained in the basic work by RR.

The question we shall examine in this section is how many factors can be said to characterize the return generating process for securities traded on the New York and American stock exchanges. This question was raised by RR who assert (Table II, p. 1088) that in about 88% of their groups (about 37 out of 42) "the probability that no more than five factors are needed to explain returns" is higher than .5, and (Table III, p. 1092) that when c_{to} is not taken as known, in 95% of the groups three or fewer factors have associated "risk premium significant at the 95% level."

As we have pointed out in section 3 one cannot test unambiguously the "significance" of individual risk premia although one can test unambiguously the null hypothesis

$$c^*_{t\cdot} = 0$$

where

$$c_{t\cdot} = (c_{to}, c^*_{t\cdot})$$

i.e., $c^*_{t\cdot}$ is the vector of risk premia. Thus, the important issue in this research is how many factors there are rather than how many "priced" factors there are. RR and others who followed their lead, such as Chen [3] among others, settle on five factors. Apparently, previous researchers, in the process of establishing "the number of factors" by the conventional likelihood ratio test (asymptotically a chi-squared test), have not appreciated the connection between how many factors are found and the number of securities one considers.[6] Table 2 presents chi-square tests on one to five factors for each of the 42 groups in the RR analytical framework. Our results differ appreciably from those obtained by RR (Table II, p. 1088). Thus, if we interpret the "probability that no more than five factors are needed to explain returns" as the p-value associated with the statistic involved in testing for a 5-factor decomposition, the comparison is as follows:

p-value	.9	.8	.7	.6	.5	.4	.3	.2	.1	0
RR	38.1	16.7	7.1	2.4	12.0	2.4	4.8	4.8	9.6	2.4
Ours	0	12.0	9.5	4.8	16.7	11.9	11.9	9.5	7.1	16.6

In the above the entry below each p-value, say .5, is to be interpreted as the percent of groups with p-values in the interval [.5,.6). Thus, for example, our results for p-value equal to zero show that seven groups (16.6%) have a p-value which indicates that a five factor decomposition is inadequate at the 10% level of significance, while RR find only 1 such group (2.4%). The difference in the findings may be attributed either to the very large number of missing observations for some securities in the RR sample or to the greater precision of our computer software (SAS) or both.[7] The differences in sample coverage between our set and that of RR amount only to 24 firms (see the description of our data).

To examine the issue of "how many factors there are" we consider the case of an expanding "universe" of securities and we give these results in Table 3. Thus, if one considers groups of, say, 15 securities, then only one or two factors may be "found;" if one considers 30 securities, then two or three may be "found." This becomes plausible if one understands the operational significance of the test, instead of concentrating solely on the purely abstract and synthetic concept of factors. What the test does (see, for example, Morrison (13), p. 269ff) is to test the hypothesis that after the extraction of k roots of the appropriate matrix the remaining roots are equal--and presumably small. Thus, looking at the 15×15 reduced correlation matrix entailed in the use of 15 security groups, it would not be surprising to find only one to two "distinct" characteristic roots; as we enlarge the scope of the investigation by dealing with, say, 30 security, 45 security, 90 security or 240 security groups, we should not be surprised if we encounter more distinct characteristic roots. This is well illustrated in Table 3 which gives the chi-squared statistics and p-values associated with a given number of factors for (overlapping) groups of 15, 30, 45, 60, and 90 security groups. We also give the p-value associated with five factors in the case of a group consisting of 240 securities.

We remind the reader that the p-value[8] is the probability that a (chi-squared) statistic at least as large as the one obtained would be realized if the null hypothesis is true, i.e. if there are at most k factors and the remaining m-k characteristic roots are the same. Choosing a level of significance at .05, we see from Table 3 that for the group of 15 securities we have at most a two-factor model; for the group of 30 securities (containing the initial 15) we have a three-factor model; for the group of 45 securities we have a four-factor model; for the group of 60 securities a six-factor model and for a group of 90 securities a nine-factor model. While these results have been obtained with a certain set of 240 securities, we have no reason to believe that, aside from singularity problems, the same phenomenon will not manifest itself with another group of 240 securities. The interesting question is at what level will the number of factors stabilize so that adding more securities to the universe will not change the number of factors conventional testing procedures will produce. We reserve this line of inquiry for another paper.

We do, however, report in Table 4 the estimates of the factor loadings from a group of 240 securities and those obtained if we factor analyzed its eight constituent groups of 30 securities each. Due to space limitations we only present this comparison for the first 60 securities. Interestingly, for any given security, estimates of the first factor loading do not change erratically as the "universe" expands. Estimates of the factor loadings for the remainder of the factors, however, change dramatically as the "universe" expands.

The import of this aspect of our work, then, is that the empirical finding in RR that the return generating process may be adequately characterized by a five factor scheme is not supported. It is then still an open question as to how many factors give an adequate characterization, but it is almost certain that there are more than five.

7. How Well Does the APT Model Explain Daily Returns?

In dealing with complex estimation procedures like those entailed by the APT model, it is not straightforward to determine just what is the explanatory power of the model or alternatively what is a measure of the "goodness of fit." We have chosen to measure "goodness of fit" by the (mean) square of the correlation coefficient between "predicted" and actual rates of return within the sample period for each group. For more details on why this is a useful measure see Dhrymes [6], chapter 2. We shall first give an account of the procedure and then discuss our findings.

We designate, in the RR methodological context, the estimator of the vector of coefficients, $c_{t\cdot}$, (involving the "riskfree" rate and risk premia) by

$$\tilde{c}_{t\cdot}^{(i)'} = (\tilde{B}_i^* \tilde{\Psi}_{ii}^{-1} \tilde{B}_i^{*'})^{-1} \tilde{B}_i^* \tilde{\Psi}_{ii}^{-1} r_{t\cdot}^{(i)'} , \qquad \begin{array}{l} i = 1, 2, \ldots, 42 \\ t = 1, 2, \ldots, T \end{array} \tag{34}$$

within each "cross section," or group of 30 securities. Thus, we have a collection of T estimates for such coefficients for each group. Owing to the fact that generalized least squares procedures are employed in the estimators of (34), the usual R^2 is not very useful or meaningful.

Thus, we use the estimates in (34) to "predict" rates of return, by

$$\tilde{r}_{t\cdot}^{(i)'} = \tilde{B}_i^{*'}(\tilde{B}_i^{*}\tilde{\Psi}_{ii}^{-1}\tilde{B}_i^{*'})^{-1}\tilde{B}_i^{*}\tilde{\Psi}_{ii}^{-1}r_{t\cdot}^{(i)'} = A_i r_{t\cdot}^{(i)'} \qquad \begin{matrix} i = 1,\ 2,\ \ldots,\ 42 \\ t = 1,\ 2,\ \ldots,\ T \end{matrix} \qquad (35)$$

Then we compute (over T observations) the square of the correlation coefficients between the actual and predicted values within each group, i.e., we compute

$$R_{ij}^2 = \left[\mathrm{Corr}\left(r_{tj}^{(i)},\ \tilde{r}_{tj}^{(i)}\right)\right]^2 \qquad j = 1,\ 2,\ \ldots,\ 30$$

The statistics given in Table 5 refer to

$$R_i^{*2} = (1/30) \sum_{j=1}^{30} R_{ij}^2 \qquad (36)$$

Consequently, what we have in Table 5 are statistics that give a measure of the mean explanatory power or goodness of fit for the 30 securities in the i^{th} group. We have done this in the case where we have used only one factor and where we have used five factors.

Several observations are in order. First, clearly at least one of the remaining four factors contributes importantly to the explanation of returns in the context of the APT model. Thus, the typical R_i^{*2} for the five factor case is about twice the corresponding R_i^{*2} in the one factor case. Second, the typical correlation is of the order of .3 in the case of five factors and about .15 in the case of one factor.

8. Additional Tests of the APT Model

We have carried out two additional tests of the APT which we cannot present in detail here in view of space limitations, but which we shall summarize in view of the importance RR place on them. These two tests consist of (1) an investigation of the relationship between the constant term or intercept in the RR cross-section regressions and the theoretically expected risk-free rate or zero-beta return,[9] and (2) an analysis of the pricing of residual variance vs. covariance measures of risk. We should note, however, that the procedures followed in the two tests, like those used by RR, are subject to the basic limitations in the application of factor analysis discussed earlier in this paper.

Our intercept analysis for the 42 groups of 30 securities each does not reject the hypothesis that the intercepts are, on average, the same in all groups, which is consistent with the APT model, but does indicate that the intercepts are insignificantly different from zero and/or significantly different from the risk-free rate for most groups, which is inconsistent with the model.[10] When standard deviation and skewness are introduced into the cross-section asset-return regressions, while generally yielding insignificant coefficients, they turn out to be significant at least as frequentely as the factors suggested by RR, which again is inconsistent with the APT model.[11]

9. Conclusions

In this paper we sought to re-evaluate the empirical evidence bearing on the relevance of the APT model; although our findings are applicable to several other studies -- many still unpublished -- we have centered our discussion on the important paper by RR which constitutes the locus classicus of this literature.

Certain important conclusions emerge from our analysis. First, the basic methodology of analyzing small groups of securities in order to gather confirmatory or contrary evidence relative to the APT model is seriously flawed. Given the theoretical foundations put forth by Ross [15], we must in our empirical analysis treat all securities symmetrically; if that is not possible because of computer software limitations, then other ways consistent with the basic requirements of the model have to be found. Analyzing small groups of securities produces results whose meaning is unclear and which cannot possibly be what the investigator wishes to accomplish.

Second, because of the indeterminacies of factor analysis, it is not possible to test directly whether a given "factor" is priced, i.e., it is not meaningful to carry out "t-tests" (or other similar tests) of significance on individual risk premia coefficients. We can, however, carry out unambiguously "F-tests" or asymptotic chi-square tests on the significance of the <u>vector of risk premia</u>. Thus, the important research issue here is how many factors there are and whether (collectively) they are priced.

Third, the basic conclusion of RR that there are three to five factors does not appear to be robust; our results show that how many factors one "discovers" depends on the size of the group of securities one deals with. For example, when dealing with a 15-security group one "discovers" two factors; when dealing with a group of 30 securities, one "discovers" three factors, with a group of 45 securities, four factors, with a group of 60 securities six factors, and with a group of 90 securities nine factors. While this exercise has not been repeated with many groups and thus we cannot put forth the proposition that we tend to "discover" factors equal to 10% of the number of securities in the group analyzed, still it raises grave doubts about the major empirical import of previous research, viz. that three to five factors give an adequate characterization of (common) market risk. This finding also suggests a possible fruitful line for future research.

Finally, we ought to point out that our findings have relevance more to the empirical methodology currently in use for testing the APT, rather than to the validity of arbitrage pricing theory models <u>per se</u>.

Footnotes

[1]Note that if $\mathrm{Cov}(f'_{t\cdot}) = \Phi$, $\Phi > 0$ otherwise arbitrary, then $f_{t\cdot}B$ is indistinguishable from $f^o_{t\cdot}B^o$ where, for arbitrary nonsingular C,

$$f^o_{t\cdot} = f_{t\cdot}C, \qquad B^o = C^{-1}B .$$

[2]See Huberman [8] for a precise definition of arbitrage and a concise yet elegant exposition of the APT model.

[3]As we point out later, at a more appropriate juncture, it is somewhat incongruous that we insist on time stationarity for the distribution of f (the risk "factors") while here we do not insist on the "stationarity," i.e. the constancy, of the risk premia c_{ti}, $i=1,2,\ldots,k$.

[4]One significant limitation that is found in all papers attempting to estimate or test the APT model is that in the estimation of the relevant covariance matrix it is assumed that the mean return process is time invariant. This is so since the (sample) covariance matrix computed by any factor analytic software package will, barring instructions to the contrary, compute the mean of the security returns as if the return generating process had a time stationary mean -- in this context, a constant mean. Actually, there is a further logical problem; in the context of the APT model, the justification for the risk premia vector is the presence of the risk factors through the term $f_{t\cdot}B$, in the return generating function. Since the "factors" are assumed to have a time stationary distribution, indeed to be independent identically distributed with mean zero and covariance matrix I, and since B is also assumed to be a matrix of fixed constants, it is difficult to understand, conceptually, why the risk premia vector should not also be time stationary. Thus, the only component of (10) for which the model carries no such implications is the constant term, c_{to}, which may be interpreted as the risk-free rate.

[5]In an earlier version of this paper, we had given a counterexample illustrating these issues; in the interest of brevity we have eliminated it from the current vesion. The counterexample, however, is available from the authors on request.

[6]Factor analytic methods have been frequently used in the finance literature. In fact, the dependency between number of factors and number of securities was noted by Meyers [12] without an explanation. See Elton and Gruber [7], Chapter 6, for a good summary of the applications of factor analysis in finance and for an extensive literature survey.

[7]We also used the same software package (EFAP) used by RR. These results are similar to those obtained by the SAS package. In their recent paper, Cho, Elton and Gruber [4] found results similar to ours.

[8]To be precise, let ξ be the test statistic, which is chi-squared with r degrees of freedom. If the level of significance is 10%, then the acceptance region is defined by

$$\text{Probability } (\xi < t_{.10}) = .9$$

where $t_{.10}$ is the boundary of the acceptance region defined by the specified level of significance. Hence

$$\text{Probability } (\xi > t_{.10}) = .1$$

both statements under H_o. The p-value that is associated with a given statistic s is then

$$\text{Probability } (\xi > s \mid H_o) = \text{p-value}$$

Hence, in order to "accept" a hypothesis at, say, the 10% level of significance, the test statistic obtained must have an associated p-value of at least .1. Of course, if the associated p-value is greater than .1, it may be the case that the hypothesis accepted contains "redundant" factors. For example, in Table 4 and for the case of 15 securities, the p-value associated with one factor is .0023, hence the one factor model should be rejected at the 5% significance level; the p-value for the two factor model is .4140, which means that this should be accepted; the p-value for the three factor model is .7676, which is also to be "accepted." On the other hand, this really contains one "redundant" factor.

[9]Under certain conditions, Ingersoll [10] argues that the intercept in the APT could be a "zero beta" asset even though a risk-free asset exists. However, this would seem to imply that the market does price risk other than common or factor risk and that arbitrage pricing theory, unlike the CAPM, cannot explain the basic risk premium between risky and risk-free assets.

[10]This finding is inconsistent even with the zero-beta version of the model since the required rate of return on such an asset is presumably in excess of zero.

[11]These additional results are available in a longer version of this paper.

References

1. S.J. Brown and M.I. Weinstein. "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm." Journal of Finance 38, (June 1983), 711-44.

2. Gary Chamberlain and Michael Rothschild. "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." Unpublished Manuscript. University of Wisconsin at Madison, June 1981.

3. Nai-Fu Chen. "Some Empirical Tests of the Theory of Arbitrage Pricing." Unpublished Manuscript. University of Chicago, Graduate School of Business, April 1982.

4. D. Chinhyring Cho, Edwin J. Elton, and Martin J. Gruber, "On the Robustness of the Roll and Ross APT Methodology," Unpublished Manuscript, Graduate School of Business Administration, New York University, 1982.

5. Phoebus J. Dhrymes. "On the Testability of Arbitrage Pricing Theory Models," Discussion Paper No. 184. Department of Economics, Columbia University, February 1983.

6. __________. Introductory Econometrics. Springer-Verlag, New York, 1978.

7. Edwin J. Elton and Martin J. Gruber. Modern Portfolio Theory and Investment Analysis. New York: John Wiley and Sons, 1981.

8. Gur Huberman. "A Simple Approach to Arbitrage Pricing Theory." Journal of Economic Theory 28 (1) (October 1982), 183-191.

9. Patricia Hughes. "A Test of the Arbitrage Pricing Theory." Unpublished Manuscript. University of British Columbia, August 1981.

10. Jonathan E. Ingersoll, Jr. "Some Results in the Theory of Arbitrage Pricing." Unpublished Manuscript. University of Chicago, Graduate School of Business, May 1981.

11. J.O. Jobson. "A Multivariate Linear Regression Test for the Arbitrage Pricing Theory." Journal of Finance 37 (September 1982), 1037-42.

12. Stephen Meyers. "A Re-examination of Market and Industry Factors in Stock Price Behavior." Journal of Finance 28 (June 1973), 695-705.

13. Donald F. Morrison. Multivariate Statistical Methods. New York, McGraw-Hill Book Company, 1967.

14. Richard Roll and Stephen A. Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." The Journal of Finance 35 (December 1980), 1073-1103.

15. Stephen A. Ross. "The Arbitrage Theory of Capital Asset Pricing." Journal of Economic Theory 13 (December 1976), 341-60.

16. __________. "Return, Risk and Arbitrage." In Irwin Friend and James L. Brickler, eds., Risk and Return in Finance. Balinger Books, Cambridge, Massachussetts, 1977.

17. Marc Reinganum. "The Arbitrage Pricing Theory: Some Empirical Results." Journal of Finance 36, (May 1981), 313-21.

18. J. Shanken. "An Analysis of the Arbitrage Pricing Theory and Its Empirical Testability." Journal of Finance 37, (December 1982), 1129-1140.

DESCRIPTION OF THE DATA

Source:	Center for Research in Security Prices Graduate School of Business University of Chicago Daily Stock Returns Files
Selection Criteria:[a]	By alphabetical order of 42 groups with the size of 30 individual securities listed on the New York and American Stock Exchanges.
Maximum Sample Size Per Security:	2619 daily returns
Minimum Sample Size Per Security:[b]	2509 daily returns
Number of Selected Securities:	1260
Time Period	July 3, 1962 to December 31, 1972

[a] Richard Roll was kind enough to give us a complete list of the companies in the RR study. We were not able to find 13 securities in the RR data in the 1982 CRSP Daily Return files due to name changes. We also replaced 11 securities which had more than 110 missing observations. In the RR sample, there are several securities with more than 800, and one with more than 1400, missing observations.

[b] The smallest sample size per group is 2424 days out of a possible 2619 days. There were 8 groups with sample size less than 2550 days, and only two groups with less than 2500 days. Note that estimation of the correlation matrix requires simultaneous observations within each group. Therefore, minimum sample size for a group can be less than the minimum size per security.

Table 1

DIAGONAL ELEMENTS OF $B_i\hat{B}_i'$ and $\tilde{B}_i\tilde{B}_i'$, i = 1,4,5,8,9,10,11,41 SHOWING LENGTH OF VECTOR OF FACTOR LOADINGS EXTRACTED FOR EACH OF THE 8 GROUPS AND LENGTH OF CORRESPONDING SUBVECTORS OBTAINED FROM THE LARGE (240 SECURITIES) GROUP

	Subgroups (30's)	Large Group Subvectors		Subgroups (30's)	Large Group Subvectors
Group 1			Group 9		
Factor			Factor		
1	3.1067	2.83768	1	3.23763	2.97218
2	.35945	.096835	2	.435946	.064010
3	.23551	.198218	3	.286408	.349995
4	.22128	.221701	4	.258647	.167212
5	.19161	.052237	5	.240895	.228194
Group 4			Group 10		
1	3.02258	2.71746	1	2.94329	2.68487
2	.24484	.171007	2	.306789	.054957
3	.25544	.159556	3	.905512	.249841
4	.22275	.155994	4	.233484	.177722
5	.18105	.122979	5	.180469	.139429
Group 5			Group 11		
1	3.54568	3.16758	1	2.54230	2.49865
2	.339056	.484800	2	.353051	.111158
3	.285417	.271622	3	.259849	.254911
4	.216376	.171041	4	.243409	.150558
5	.202698	.093324	5	.212725	.184807
Group 8			Group 41		
1	3.24736	2.93784	1	3.33370	2.93098
2	.343372	.051216	2	.330169	.460566
3	.268385	.249631	3	.250717	.261120
4	.192920	.146789	4	.233161	.115732
5	.196137	.136052	5	.201804	.075048

In view of Equation (26), corresponding entries in the two columns should be identical, allowing for round-off error. A departure from this identity is not a matter of statistical significance, and its occurrence implies that Equation (26) is false. The eight groups are numbered following the RR classification scheme.

Table 2

CHI-SQUARE TEST OF THE HYPOTHESIS THAT k-FACTORS GENERATE DAILY SECURITY RETURNS FOR SELECTED GROUPS

GROUP	k=1	k=2	k=3	k=4	k=5
1	572.35 (.0001)	435.41 (.0185)	372.63 (.1742)	318.38 (.5209)	269.37 (.8554)
6	873.95 (.0001)	659.04 (.0001)	452.08 (.0001)	362.26 (.0560)	309.89 (.2643)
11	611.09 (.0001)	472.36 (.0005)	399.48 (.0295)	333.67 (.3016)	280.32 (.7214)
16	570.17 (.0001)	455.25 (.0031)	364.93 (.2556)	312.95 (.6157)	273.87 (.8062)
21	692.12 (.0001)	596.79 (.0001)	502.53 (.0001)	444.129 (.0001)	382.87 (.0004)
26	547.99 (.0001)	441.82 (.0107)	377.27 (.1304)	315.79 (.5716)	277.64 (.7586)
31	553.53 (.0001)	466.29 (.0010)	394.49 (.0433)	327.30 (.3923)	288.28 (.5592)
36	606.24 (.0001)	493.53 (.0001)	426.06 (.0027)	369.29 (.0326	314.41 (.2091)
42	565.53 (.0001)	457.46 (.0025)	404.03 (.0205)	344.85 (.1723)	300.89 (.3942)
Number of Times the Null Hypothesis is Accepted at 5% Significance Level for the Entire 42 Groups					
	0	1	11	29	36

Number of degrees of freedom for the chi-square test for each group is $\frac{1}{2}[(30-k)^2 - 30 - k]$. Number of securities in each group is 30. Number of observations for each security is 2618. No security has more than 110 missing observations. Values in parentheses indicate the p-value associated with the statistic, i.e., the probability that the test statistic (under the null hypothesis) will assume a value at least as large as the statistic obtained in this particular test. Groups are numbered following the RR classification scheme and only results for every 5th group are presented here in the interest of brevity.

Table 3

CHI-SQUARE TEST OF THE HYPOTHESIS THAT k-FACTORS GENERATE DAILY SECURITY RETURNS USING VARYING GROUP SIZES

Number of Factors (k)	Number of Securities in a Group 15	30	45	60	90	240
1	132.6 (.0023)	572.4 (.0001)	1246.4 (.0001)	2318.7 (.0001)	5548.9 (.0001)	
2	78.0 (.4140)	435.4 (.0189)	1065.1 (.0001)	2057.9 (.0001)	4986.59 (.0001)	
3	54.5 (.7676)	372.6 (.1742)	958.41 (.0094)	1845.6 (.0001)	4501.7 (.0001)	
4	37.7 (.9165)	318.4 (.5309)	858.6 (.1461)	1697.4 (.0023)	4190.5 (.0001)	
5	28.5[a] (.9132)	269.4 (.6554)	776.5 (.4785)	1603.3 (.0133)	3962.8 (.0001)	30756.2[b] (.0001)
6		230.3 (.9617)	711.2 (.7290)	1502.1 (.0762)	3776.7 (.0003)	
7		199.3 (.9869)	658.9 (.8403)	1409.4 (.2301)	3606.3 (.0061)	
8		165.6 (.9985)	610.6 (.9067)	1320.7 (.4735)	3460.0 (.0369)	
9		139.7 (.9999)	558.3 (.9660)	1247.4 (.6377)	3321.7 (.1299)	
10		117.4 (.9999)	570.4 (.9882)	1159.88 (.8691)	3188.9 (.3091)	

Figures in the first line are the chi-squared values. Figures in the paren##omputed as $\frac{1}{2}[(n-k)^2 - n - k]$, where n is the number of securities in the group and k is the number of factors. Only five factors are estimated for the group of 240 securities.

[a]It is not possible to extract more than 5 factors.

[b]Only five factors are estimated due to accelerating computer costs.

Table 4: EFFECT OF GROUP SIZE ON THE ESTIMATES OF FACTOR LOADING

SECURITY	FACTOR 1		FACTOR 2		FACTOR 3		FACTOR 4		FACTOR 5	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
1	0.34301	0.35724	0.01151	-0.11849	0.08339	-0.15421	-0.10169	-0.09776	0.08887	0.02068
2	0.26761	0.29088	0.01657	0.14466	-0.01217	0.11512	0.14923	-0.06598	-0.02374	-0.04764
3	0.45764	0.45577	0.05844	-0.03549	-0.09902	0.07795	-0.05537	0.00672	0.03867	0.00237
4	0.48656	0.49847	0.09482	-0.15348	-0.21406	0.14236	-0.01419	-0.01614	-0.04208	0.07065
5	0.26938	0.29360	0.01804	0.06250	0.03573	-0.00284	0.03926	-0.03192	0.00914	0.11346
6	0.25149	0.31159	0.07037	0.19731	0.07316	-0.20100	0.07671	0.00950	0.04320	0.08345
7	0.17076	0.19279	-0.00823	0.06214	0.09163	-0.10155	0.00381	0.06930	0.02038	-0.05445
8	0.29154	0.30195	0.03549	-0.03405	0.05869	0.01576	-0.05353	0.08203	-0.00401	0.01626
9	0.14089	0.16681	0.05705	0.14805	0.03437	0.00504	0.14623	-0.00019	0.03386	0.04758
10	0.23886	0.25228	-0.01933	-0.05308	0.07894	-0.05234	-0.05091	0.03391	0.06375	0.07454
11	0.10899	0.14624	0.03967	0.15365	0.07481	-0.01931	0.09424	0.05798	0.00760	-0.03799
12	0.16868	0.18307	0.00971	0.11081	0.07766	0.00805	0.07895	0.00955	0.03468	-0.05260
13	0.28098	0.28760	0.01658	0.02965	0.12869	-0.09825	-0.07218	-0.06310	-0.00477	-0.10965
14	0.34900	0.35310	0.06631	0.09803	0.01324	0.03121	0.12312	-0.00881	0.01049	0.15848
15	0.46722	0.47271	0.05190	-0.03094	-0.15800	0.04911	-0.08409	-0.11453	0.05553	0.06990
16	0.22706	0.27374	0.05516	0.16885	0.00954	-0.02110	0.10702	0.02568	-0.05370	0.01330
17	0.21684	0.26351	0.06742	0.17886	0.05731	0.03293	0.13076	0.01768	0.01336	-0.02118
18	0.26804	0.28819	0.01419	0.01000	-0.02847	0.13687	0.10407	0.12255	0.04402	0.10326
19	0.20803	0.19866	0.01086	0.07676	0.05684	-0.03509	0.06085	0.07363	-0.03545	-0.00995
20	0.38283	0.39971	-0.00874	-0.13148	0.01843	-0.12364	-0.12410	0.05489	0.04228	0.04205
21	0.45467	0.44895	-0.01909	0.00432	0.07669	-0.04804	-0.01463	-0.03987	0.04025	-0.04668
22	0.34498	0.35191	-0.01179	-0.08561	0.02781	-0.02926	-0.08141	0.04670	0.03636	-0.04906
23	0.12078	0.13405	0.02868	0.07965	0.11959	-0.13553	-0.06627	-0.22166	0.02321	0.05291
24	0.20618	0.20640	-0.22441	-0.02853	0.04288	-0.11423	0.02303	0.07416	0.06892	-0.05730
25	0.25661	0.26771	0.01285	0.07071	-0.01387	0.12422	0.07418	-0.19787	0.03214	-0.04322
26	0.50319	0.51228	-0.05831	-0.14862	-0.07452	0.05006	-0.11978	-0.05966	0.01733	-0.10228
27	0.22149	0.24849	-0.02304	0.10345	0.03685	0.06609	0.09502	0.16071	0.06919	0.10227
28	0.39352	0.41647	-0.00137	0.03893	0.03112	0.02107	0.04287	0.06735	-0.05720	-0.17248
29	0.25206	0.25412	-0.03066	0.09717	0.13312	0.04712	0.04144	0.06267	-0.05256	-0.17396
30	0.29395	0.32082	0.00588	-0.16542	-0.00557	-0.09269	-0.07941	0.11322	0.00312	0.01288

Table 4 (continued)

SECURITY	FACTOR 1 (a)	FACTOR 1 (b)	FACTOR 2 (a)	FACTOR 2 (b)	FACTOR 3 (a)	FACTOR 3 (b)	FACTOR 4 (a)	FACTOR 4 (b)	FACTOR 5 (a)	FACTOR 5 (b)
31	0.21205	0.21198	0.07090	0.01332	0.06219	0.00521	−0.02153	−0.01223	0.00524	0.05128
32	0.23807	0.25915	0.02371	0.07730	0.03052	0.16568	0.05748	−0.13780	0.01272	−0.07643
33	0.26478	0.29668	0.00696	0.18634	0.10602	−0.05850	−0.06638	−0.18642	−0.12960	0.04983
34	0.28912	0.31687	−0.03304	0.14878	0.10592	0.03991	−0.07752	−0.09832	−0.09863	0.02598
35	0.23751	0.27244	0.04911	0.09158	0.04311	0.16671	0.04915	0.13117	0.07946	0.04873
36	0.26254	0.27919	0.03167	0.02980	0.00398	0.05356	0.09937	−0.02001	−0.01669	0.05039
37	0.41241	0.40702	−0.36521	0.08224	−0.00582	−0.08784	0.07546	0.03160	−0.04735	0.06807
38	0.30784	0.29731	0.03028	−0.01583	0.08926	0.01638	−0.03049	0.01053	−0.00577	0.07849
39	0.32584	0.34266	0.02310	0.03574	0.00772	−0.21601	−0.07560	0.00520	−0.09633	0.00079
40	0.23206	0.26257	0.01510	0.11503	0.09286	0.09603	0.02584	0.03050	0.07580	−0.00539
41	0.21193	0.23441	−0.00166	−0.04631	0.07918	0.04014	0.10253	0.03320	0.00698	−0.13417
42	0.34305	0.37291	−0.01505	−0.00766	0.06114	0.07163	0.02607	−0.06447	0.04243	0.11717
43	0.24401	0.24909	0.04142	0.00853	−0.02423	−0.01114	0.13276	0.20230	−0.11286	0.04653
44	0.39762	0.44041	0.05887	0.04047	0.02390	−0.07102	0.04315	0.05332	−0.01232	−0.08036
45	0.28283	0.30192	0.03636	0.09511	0.11662	−0.03135	−0.11473	0.07821	−0.02905	0.06770
46	0.13576	0.12319	−0.03679	0.05422	0.03815	0.01119	0.09924	0.01295	−0.00831	−0.02105
47	0.27812	0.27343	0.01773	−0.06769	−0.00982	0.10950	0.08100	0.05530	0.02230	−0.05773
48	0.16851	0.19681	0.02790	0.02431	0.03237	0.12862	0.04847	0.02383	0.03735	−0.03807
49	0.22462	0.22965	0.00639	0.05360	−0.08456	0.02845	−0.08162	−0.09090	−0.07664	0.07176
50	0.35168	0.39336	−0.00676	0.13158	0.03084	−0.18638	0.06484	0.07740	−0.16676	−0.09556
51	0.19512	0.21469	0.01553	−0.02108	−0.06556	0.15959	−0.00852	0.03851	0.12456	−0.09492
52	0.20396	0.24635	0.03239	−0.03913	0.07650	0.09956	0.10694	−0.07547	0.00595	0.01579
53	0.50936	0.51384	0.05390	−0.20767	−0.08883	−0.04232	0.06991	0.05945	−0.02376	0.06582
54	0.51235	0.51515	0.07911	−0.21992	−0.17779	−0.01904	0.07804	−0.07762	−0.02102	0.03670
55	0.14234	0.14788	−0.02987	−0.03820	0.05303	0.03175	0.07368	−0.01377	−0.01225	0.11177
56	0.40374	0.39379	0.01991	−0.05442	−0.00689	−0.07360	−0.04476	−0.11685	0.00627	−0.02960
57	0.32436	0.30144	0.04618	−0.05421	−0.02671	−0.01982	0.00055	0.02976	−0.02102	−0.14782
58	0.40243	0.45506	0.02783	−0.02162	0.07134	0.01922	−0.00680	−0.07177	0.04908	−0.16761
59	0.22556	0.24618	0.02527	−0.02865	0.05926	0.02563	0.08455	0.03755	0.00653	0.10351
60	0.22458	0.28366	0.03544	0.08757	0.12825	0.07710	0.07703	0.16824	0.03885	0.01669

Column (a) contains the estimates of factor loadings for the kth factor k = 1, 2, ..., 5, from a group of 240 securities. Column (b) contains the estimates of the corresponding factors from groups of 30 securities, i.e., first 30 rows are the estimates from the first group; second 30 rows are the estimates from the second group, etc. Only the estimates of the factor loadings for the first sixty securities are shown here.

Table 5: SUMMARY STATISTICS FOR THE SQUARED CORRELATION COEFFICIENT BETWEEN REALIZED DAILY RETURNS AND THE FORECASTS BY ONE- AND FIVE-FACTOR MODELS

Group No.	1-Factor Model Mean	1-Factor Model Std. Dev.	5-Factor Model Mean	5-Factor Model Std. Dev.
1	.1626	.0933	.3069	.1381
2	.1802	.0859	.3184	.1833
3	.1588	.0809	.2843	.2614
4	.1617	.0952	.3013	.1290
5	.1789	.1048	.3135	.1650
6	.1536	.0681	.2933	.2164
7	.1524	.0794	.2960	.1813
8	.1734	.0695	.3105	.1259
9	.1711	.0925	.3130	.1581
10	.1619	.0957	.2994	.1483
11	.1482	.0811	.2919	.1329
12	.1649	.0816	.3034	.1740
13	.1518	.0988	.2883	.1810
14	.1637	.0791	.3033	.1493
15	.1695	.1236	.3035	.1753
16	.1768	.0913	.3126	.1621
17	.1903	.1076	.2368	.1453
18	.1675	.1023	.2991	.1983
19	.1540	.0905	.2928	.1515
20	.1529	.0774	.2936	.2423
21	.1685	.0814	.3052	.1702
22	.1712	.0828	.3081	.1373
23	.1533	.0852	.2914	.1754
24	.1693	.0911	.3018	.1842
25	.1672	.0793	.3056	.1677
26	.1612	.0690	.3026	.1007
27	.1524	.0897	.2957	.1458
28	.1639	.0829	.2959	.1894
29	.1508	.1005	.2899	.1695
30	.1547	.0791	.2959	.1462
31	.1560	.0824	.2943	.1678
32	.1496	.0749	.2971	.1571
33	.1617	.0841	.2970	.1856
34	.1459	.0859	.2882	.1728
35	.1655	.0872	.3042	.1480
36	.1580	.0866	.2880	.2190
37	.1588	.0973	.2957	.1731
38	.1460	.0684	.2842	.2041
39	.1467	.0822	.3019	.1832
40	.1828	.1088	.3212	.1830
41	.1720	.1050	.3091	.1634
42	.1486	.0825	.2879	.1962

Forecasts of daily returns are estimated by $\tilde{r}'_{t.} = \tilde{B}^{*\prime}(\tilde{B}^{*}\tilde{\Psi}^{-1}\tilde{B}^{*\prime})^{-1}\tilde{B}^{*}\tilde{\Psi}^{-1}r'_{t.}$. See equation (35) for definition of variables. Mean is the average of squared correlations between the forecast, $\tilde{r}_{t.}$, and the realized daily return, $r_{t.}$, for 30 securities in each group. Std. Dev. is the standard deviation of the squared correlation for the 30 securities in each group.