The Number of Factors in Security Returns
Author(s): Stephen J. Brown
Source: *The Journal of Finance*, Vol. 44, No. 5 (Dec., 1989), pp. 1247-1262
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2328641
Accessed: 26/11/2009 07:31

THE JOURNAL OF FINANCE • VOL. XLIV, NO. 5 • DECEMBER 1989

The Number of Factors in Security Returns

STEPHEN J. BROWN*

ABSTRACT

Both factor analysis of security returns and the analysis of eigenvalues seem to indicate that a market factor explains the major part of security returns. We find that such evidence is consistent with an economy where there are in fact k "equally important" priced factors; eigenvalue analysis in the context of such an economy will lead an investigator to the false inference that the one important "factor" is the return on an *equally weighted* market index.

THIS PAPER EXAMINES A particular "exact" k-factor economy where each of the k factors is priced and contributes equally on average to the variance of returns. We show analytically that for such an economy the approximate factor solution identifies the first factor as the return on an equally weighted "market" index and would lead to the inference that the remaining $k - 1$ factors are not priced using a cross-section regression technology. Our particular model is consistent with existing empirical work that has studied the number of factors in security markets. These results suggest an important caveat to the interpretation of approximate factor solutions in an APT or multibeta pricing context. One might also infer a CAPM bias in tests of the APT using the factor analysis technology. On a more positive note, the results suggest a simple rotation of the factor analysis solution that will lead to more powerful tests of the asset pricing relations.

To motivate this study, consider the fact that the single-factor market model suggested by Markowitz (1959) and developed by Sharpe (1963) remains one of the most popular empirical models of the return-generating process. The widespread acceptance of this model owes much to the work of King (1966). Through a seven-factor analysis of return data, he concludes that one factor explains a large percentage of the variance of stock prices, a factor on which each security tends to weight positively. He interpreted this result to mean that there exists a basic market factor in operation which has a major effect on all securities. This work supported the development of empirical research into the Capital Asset Pricing Model. The model has been extensively applied in the event study context

* Yamaichi Faculty Fellow, Leonard N. Stern School of Business, New York University. I wish to acknowledge helpful comments of Michael Brennan, Gary Chamberlain, Dolores Conway, Phil Dybvig, Ned Elton, David Fowler, Jon Ingersoll, Shmuel Kandel, Steve Ross, Mike Rothschild, Ravi Shukla, Charles Trzcinka, and especially Jay Shanken, David Mayers (the co-editor), and the referee, as well as the participants of the Empirical Research in the Arbitrage Pricing Theory conference held at the University of Southern California, November 1985. I do not implicate any of these people in the errors that remain.

(Brown and Warner (1980, 1985)), and work by Brown and Weinstein (1985) shows that the use of an equally weighted market index as a single factor would appear to do almost as well as a four-factor representation of security returns in this application.

The issue of determining the appropriate number of factors is, of course, central to the Arbitrage Pricing Theory (APT) of Ross (1976), where a factor structure was assumed to generate observed security returns. In this context factors are interpreted as fundamental sources of nondiversifiable risk. Roll and Ross (1980) re-examine the issue of the number of factors by studying the extent to which factors generated by a factor analysis were priced in the capital markets and could thus be considered nondiversifiable. They find a five-factor representation adequate for their purposes. Using likelihood ratio procedures, Brown and Weinstein (1983) find that the number of factors common across securities was between three and five.

Factor analysis in this rate of return context suffers from two distinct limitations. The first is that the data may violate the assumptions of the model. In particular, this analysis requires an *exact factor structure* in which the residual component of returns not explained by the factors is uncorrelated across securities. The second problem is that the factor analysis factors are unique only up to rotation. The fact that there is a certain indeterminacy associated with this lack of identification is the chief barrier against finding a significant economic interpretation to the results of factor analysis in this context.

In a very important contribution, Chamberlain and Rothschild (1983) (CR) propose an alternative procedure that addresses these issues. Instead of assuming an exact factor structure, to be analyzed through factor analysis or other similar procedures, they assume an *approximate factor structure* in which residual returns may be weakly correlated across securities. For such an approximate k-factor structure, the first k eigenvalues of the covariance matrix of returns grow without limit as the number of securities, m, increases, while the remaining $m - k$ stay constant. Assuming such a factor structure, they obtain results similar to those of Ross (1976).[1]

The analysis of approximate factor structures has important empirical implications. The principal factor solution associated with the first k eigenvalues (under certain conditions) is a consistent[2] estimator of the k-factor structure. A major argument in favor of the principal components procedure is that the estimator derived in this manner is computationally and conceptually simpler than factor analysis. The approximate factor structure is unique. The analysis seems to address the rotation problem often associated with the use of factor analysis. Furthermore, since the factor structure determines the way that the principal components grow as the number of securities increases, CR (p. 1285) suggest an important role for principal components in determining the number of factors.

[1] Similar results are obtained by Ingersoll (1984) under somewhat weaker conditions.

[2] In this context, it is important to note that the results we report in this paper apply for finite m (number of securities) and large t (number of observations). Hence, whenever in this paper we use the term "consistent," we mean consistent as t grows large. Shanken (1985) also draws the important distinction between m-consistency and t-consistency.

Trzcinka (1986) studies this empirical insight. On the basis of 1069 weeks of return data on 865 securities, he computes sample covariance matrices of successively higher dimension. He then extracts the largest eigenvalues. The first eigenvalue contributes a large fraction of the variance of returns, and its relative importance increases dramatically with the number of securities. While this observation does not constitute a formal statistical test, it would suggest that there is one major factor generating security returns. The remaining eigenvalues also appear to increase with the number of securities, albeit at a smaller rate, and it is otherwise impossible to distinguish among these eigenvalues. One concludes from this study that, while there may be many factors responsible for security returns, one factor appears more important than the rest. Luedecke (1984) finds similar results. Using slightly different procedures, Linn and Chang (1985) find one large eigenvalue and possibly a second smaller one as the number of securities increases.

These results have a significance that extends beyond the study of approximate factor structures. The principal factor solution is used as the starting value in maximum-likelihood factor analysis algorithms. To the extent that the principal factor solution is a consistent estimator of the underlying covariance structure of returns, for a sufficient number of time series observations this solution will satisfy necessary conditions for a maximum to the factor analysis likelihood function. Thus, the findings of Trzcinka, Luedecke, and others have relevance for the results others have found using the method of factor analysis.

The existing studies of approximate factor structures rely on fairly informal observations on the pattern of ranked eigenvalues as the number of securities increases. Any formal statistical test for the number of factors in this context would depend on little understood small sample properties of ranked eigenvalues. In addition, other than the recent work of Connor and Korajczyk (1988), little effort has been expended to discover whether the factors implied by the approximate factor structure are priced. In their work, Roll and Ross (1980) suggest that the uniqueness problem associated with factor analysis is not an issue since the APT results are invariant to rotation of the original factors. While this statement is true, Chen (1983) points out that statistical tests for the number of priced factors are not invariant to rotation. The approximate factor structure solution is unique. It is important to discover the effect the particular rotation implied by this solution has on the tests of asset pricing.

This paper is in two parts. The first part presents the simple economy and describes the relationship between eigenvalues and the number of underlying securities as the number of securities increases in such an economy. It describes the nature of the approximate factor solution and implications for factor pricing. This simple model has a strong prediction on the role and importance of the equally weighted market index. This section concludes with an examination of the empirical evidence on this point. These results assume that the investigator has access to the true parameters of the factor structure of returns. This assumption is relaxed in the second part of the paper. There we examine the effect of having only a limited number of time series observations to estimate the factor structure. We study the small sample properties of ranked eigenvalues using data simulated to have come from the simple economy introduced in the first part of the paper.

I. Analytical Findings

In this section we present results for a simple economy for which we can derive eigenvalues analytically. We show how the eigenvalues grow with the number of securities and how the implied approximate factor solution for this economy relates to the original factor structure. We are then in a position to consider the effect such an approximate factor solution has on tests of factor pricing. We conclude with some observations on the extent to which an important implication of the model seems to be borne out in the empirical evidence.

A. The Model

Consider an economy where the k factor loadings for a randomly selected security can be considered realizations from a normal distribution with mean one and $k \times k$ covariance matrix[3] $\sigma_b^2 I_k$. The factor loadings for m securities can be represented as an $m \times k$ matrix B, $B = \{b_{ij}\}$. Given this matrix, the assumption that the factors are independent and identically distributed with variance σ_f^2, and the $m \times m$ covariance matrix of idiosyncratic components $\sigma_e^2 I_m$, the eigenvalues for the covariance matrix of returns

$$\Sigma_R = BB'\sigma_f^2 + \sigma_e^2 I_m$$

can be derived[4] for large m:

$$\lambda_1 = \sigma_e^2 \left[\frac{R^2}{(1 - R^2)k} [(m - 1)\sigma_b^2 + k \cdot m] + 1\right],$$

$$\lambda_{2,\cdots,k} = \sigma_e^2 \left[\frac{R^2}{(1 - R^2)k} \cdot (m - 1)\sigma_b^2 + 1\right],$$

$$\lambda_{k+1,\cdots,m} = \sigma_e^2,$$

[3] This simple economy is not ruled out by the Arbitrage Pricing Theory or other equilibrum pricing models and for this reason represents a "reasonable" alternative by which to judge the power of procedures to determine the number of factors. The model is not as restrictive as it appears. In theory, the factor loadings could be drawn from a distribution with mean zero. However, this would have the unreasonable implication that an equally weighted market index would have zero variance as the number of securities increased without limit. The magnitude of the factor loadings is otherwise unique only up to scale. Furthermore, the cross-sectional covariance of the factor loadings is given only up to rotation of the original factors. The most restrictive feature of the model is the assumption that the cross-sectional variance of the factor loadings is constant across factors.

[4] Note that

$$|\sigma_f^2 BB' + \sigma_e^2 I_m - \lambda I_m| = 0$$

implies that

$$|BB' - \lambda^* I_m| = 0,$$

where λ^*, the eigenvalues of BB^*, are given by

$$\lambda_j^* = \frac{\lambda_j - \sigma_e^2}{\sigma_f^2}.$$

Since the nonzero eigenvalues of BB' are equal to those of $B'B$ and since

$$B'B = (m - 1)\, \hat{\Sigma}_b + m\mathbf{b}\mathbf{b}' = (m - 1)\sigma_b^2 I_k + m\iota\iota',$$

for large m (where we replace sample values with population counterparts and where $\mathbf{b}$ is a $k \times 1$ vector of cross-sectional mean factor loadings), which is an intraclass structure, the result follows. (See, e.g., Press (1972), p. 29.)

where we define R^2 as

$$R^2 \equiv \frac{k\sigma_f^2}{k\sigma_f^2 + \sigma_e^2},$$

interpreted as the percentage of variance explained by the factor model. The first derivatives are

$$\frac{d\lambda_1}{dm} = \sigma_e^2 \left[\frac{R^2}{(1 - R^2)k} \cdot [\sigma_b^2 + k] \right],$$

$$\frac{d\lambda_{2,\ldots,k}}{dm} = \sigma_e^2 \left[\frac{R^2}{(1 - R^2)k} \cdot \sigma_b^2 \right],$$

$$\frac{d\lambda_{k+1,\ldots,m}}{dm} = 0.$$

As pointed out by CR, the first k eigenvalues increase without bound as the number of securities increases, while the remaining $m - k$ are constant. However, the first eigenvalue is larger than the others. Furthermore, that eigenvalue increases more with the number of securities than do the others. The relative importance of the first eigenvalue depends *solely* on the ratio of the number of factors, k, to the variance of the factor loadings across securities, σ_b^2.[5] This yields the interesting result that the first eigenvalue becomes large relative to the others as the number of underlying factors increases. In other words, the approximate factor solution would imply that the larger the number of equally important factors, other things equal, the more certain would a casual empirical investigator be that only one factor is in fact responsible for security returns!

One should note that it is possible, at least in principle, to distinguish the second through kth eigenvalues from the remaining eigenvalues by virtue of the fact that those eigenvalues have a positive gradient with respect to the number of securities. Whether this is true in practice depends, of course, on the numerical accuracy of the eigenvalue computations and the sampling properties of those numbers computed on the basis of sample covariance matrices. In the next section we illustrate some of the numerical and sampling problems associated with the approximate factor solution.

B. Relationship Between Approximate Factor Solution and Underlying Factors

The first k principal factors (associated with the first k eigenvalues of the covariance matrix) for this example represent a particular orthogonal rotation of

[5] To see this, note that the ratio of the first two derivatives is given by

$$\frac{d\lambda_1}{dm} \bigg/ \frac{d\lambda_2}{dm} = 1 + \frac{k}{\sigma_b^2}$$

and that a sufficient condition for the difference between the first two eigenvalues to be greater than the difference between the kth and the $k + 1$st eigenvalues is that k be greater than σ_b^2.

the k original factors.[6] The first principal factor has the interpretation of being the return on an equally weighted market index (in excess of the mean return on that index). The second and remaining principal factors explain the variance of security returns not otherwise accounted for by returns on the index, and so it is not surprising that the first principal factor explains a large fraction of the variance of security returns. To show that the first principal factor is indeed the equally weighted market, simply note that the first principal factor loadings average to a constant different from zero while the remaining loadings average to zero in the cross-section of securities.[7] Since the cross-sectional average of the idiosyncratic terms is close to zero at any point in time, the result follows.

C. The Effect of the Approximate Factor Solution on Pricing Inferences

To this point, we have examined the purely statistical properties of the approximate factor solution in this context. It is interesting also to consider the implications for pricing inferences. Chen (1983) points out that, if it has been established that k factors are present, the number of priced factors can be any number between one and k. In the present instance, suppose that a generalized least squares regression of mean returns against the underlying factor loadings revealed that each of the original k factors were priced (with t-values equally significant and greater than two). Then it is possible to show that the first principal factor, the equally weighted market index, will appear to be a priced factor (t-value greater than $2\sqrt{k}$) and that the remaining factors implied by the principal factor solution will appear not to be priced (t-value equal to zero)[8]

[6] For each of the first k eigenvalues λ_j $j = 1, k$, $\mathbf{w}_j$ solves $[BB'\sigma_f^2 + \sigma_m^2]\ \mathbf{w}_j = \lambda_j\ \mathbf{w}_j$, where $\mathbf{w}_j$ is the jth column of B postmultiplied by the Helmert matrix, A:

$$A = \begin{bmatrix} 1/\sqrt{k} & 1/\sqrt{2} & 1/\sqrt{2\times 3} & \cdots & 1/\sqrt{(k-1)k} \\ 1/\sqrt{k} & -1/\sqrt{2} & 1/\sqrt{2\times 3} & \cdots & 1/\sqrt{(k-1)k} \\ 1/\sqrt{k} & 0 & -2/\sqrt{2\times 3} & \cdots & 1/\sqrt{(k-1)k} \\ 1/\sqrt{k} & 0 & 0 & \cdots & 1/\sqrt{(k-1)k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1/\sqrt{k} & 0 & 0 & & -(k-1)/\sqrt{(k-1)k} \end{bmatrix}$$

It should be noted that this rotation is unique only up to permutation of the original factors. The jth principal factor loading is then given by $\mathbf{w}_j$ normalized to be of unit length, multiplied by the square root of the jth eigenvalue.

[7] It is apparent from inspection of the Helmert rotation of the previous footnote that the first principal factor loading is proportional to an equally weighted average of the original factor loadings, and the second and remaining principal factor loadings are proportional to differences and differences from means of the original factor loadings. Thus, the cross-sectional average of the first and remaining principal factor loadings will be proportional to one and zero, respectively.

[8] This result follows from the special structure of Σ and the approximation to $B'B$ for large m given in footnote 6. To abstract from measurement error issues we assume that B is known. The generalized least squares estimate $\hat{\gamma}$ $[B'\Sigma^{-1}B]^{-1}B'\Sigma^{-1}\mathbf{r}$ has covariance matrix var $(\hat{\gamma}) = [B'\Sigma^{-1}B]^{-1}$, where Σ is the covariance matrix of returns corresponding to the underlying factor model and $\mathbf{r}$ is the vector of mean excess returns. Suppose instead that the investigator regressed average returns against the approximate factor solution given as $B^* = BWS$, where W is the Helmert rotation and S is a strictly positive diagonal scaling matrix, yielding $\hat{\gamma}^* = S^{-1}W'\hat{\gamma}$, and var $(\hat{\gamma}^*) = S^{-1}W'\text{var}(\hat{\gamma})WS^{-1}$. If each element of $\hat{\gamma}$ is equally significant (in terms of difference from zero) with $t(\gamma_i) = \hat{\gamma}_i/\sqrt{\text{var}(\hat{\gamma}_i)}$

regardless of whether or not the equally weighted market index is mean-variance efficient.[9]

In other words, the principal factor solution represents a rotation of the original factors that minimizes the apparent number of priced factors. These results seem to suggest a CAPM bias in tests of the APT or multibeta pricing relations which use the first k principal factors. Estimated factor structures where the first factor is identified with the equally weighted market may only be a rotation away from a factor structure with k equally important factors. However, the factor price associated with that first factor (which contributes most to the variance of returns) will be estimated more precisely than the remaining factor prices. Hence, it may be difficult to distinguish statistically between the APT and the CAPM even where the equally weighted index is not mean-variance efficient. One possible resolution of the problem is to consider the pricing of statistical factors defined as a rotation of the estimated factors, where the rotation is given as the inverse of the Helmert rotation. (See footnote 6.) This inverse rotation designed to retrieve the k "equally important" factors obviously will not affect the APT relationships. However, it will lead to more powerful tests of the asset pricing relationships to the extent that each of the factor prices would be estimated with equal statistical precision.

$= \delta$, $\delta > 2$, $i = 1, k$, what is the significance of $\hat{\gamma}^*$? First note that the covariance matrix of the original factor price estimates is idempotent:

$$\operatorname{var}(\hat{\gamma}) = [B'\Sigma^{-1}B]^{-1} \approx \kappa_1 I_k - \kappa_2 \iota\iota', \; \kappa_1 > k\cdot\kappa_2 > 0$$

$(B'\Sigma^{-1}B = B'[D^{-1} - \sigma_f^2 D^{-1}BAB'D^{-1}]B = X - \sigma_f^2 XAX = XA$, where $A = [I_k + \sigma_f^2 X]$ and $X = B'D^{-1}B \approx [(m-1)\sigma_b^2 I_k + m\iota\iota']/\sigma_e^2)$. Then the result follows directly, with the t-ratio for the estimated price of the first approximate factor greater than twice the square root of the number of factors, and the t-ratios for the remaining factor prices equal to zero:

$$t(\hat{\gamma}_1^*) = \delta\sqrt{k} \cdot \sqrt{\{(\kappa_1 - \kappa_2)/(\kappa_1 - k\kappa_2)\}} > 2\sqrt{k};\; t(\hat{\gamma}_j^*) = 0,\; j = 2, \ldots, k.$$

[9] This raises the interesting issue of whether an *equally weighted* index of the m securities under study is in fact mean-variance efficient. In general, it will not be efficient where the securities under study form a subset of the market portfolio. Most empirical studies of factor structures are confined to a subset of equities trading on the major exchanges. Some results can be obtained where there is considered an identity between the securities under study and the market portfolio and the further assumption is made that the APT holds *exactly*. Under these circumstances, Tiemann's (1988) Proposition 1 has demonstrated conditions sufficient for the equally weighted index to be mean-variance efficient. In the present instance with equal idiosyncratic risk, they will be satisfied if the APT factor prices are identical. It is an interesting exercise to derive the efficient set constants (see, e.g., Roll (1977)) for this example using the approximation for $B'\Sigma^{-1}B$ cited in the last footnote. Then, using the efficient set mathematics of Roll (1985), it is possible to show that Shanken's (1985) *CSR* T^2 metric for the equally weighted market portfolio approaches zero as the number of securities m increases without limit. However, a caveat is in order. This exercise in efficient set mathematics also shows that under these conditions the variance-minimizing portfolio with mean return equal to that of the equally weighted index has value weights that deviate from equality. The extent of the deviation from equality depends on the magnitude of σ_b^2. This paradox represents another manifestation of the Shanken (1982) puzzle. As the number of securities m becomes unbounded, the APT holds exactly but the covariance matrix Σ has no inverse and the efficient set constants are not well defined.

D. Empirical Evidence

To this point, it has sufficed to show that there exists a simple counterexample to the proposition that the approximate factor solution will correctly identify the number of economically meaningful factors. There is some evidence that this model is also empirically significant. Connor and Korajczyk (1988) find that the first principal factor generally explains over 99 percent of the variance of the equally weighted portfolio. This result is consistent with those reported by King (1966). However, Connor and Korajczyk also find the results are quite different for the value weighted portfolio, where the first factor only explains from 65 to 85 percent of the variance of that portfolio. As further evidence, Brown and Weinstein (1985) find that a four-factor representation of security returns appears to contribute little value added over the use of an equally weighted market index in the context of event studies. The same result was found by Chen, Copeland, and Mayers (1987) in the context of performance measurement studies.[10] In fact, most empirical work seems to identify the first factor as an equally weighted market index.

This empirical evidence is obviously consistent with the existence of a pervasive market factor. The point of the example presented in this paper is that the example is *also* consistent with k equally important factors that are priced. Thus, it would be misleading to conclude from such empirical results alone the existence of a pervasive market factor; it would be simply wrong to conclude that such results justify the use of a *value weighted* market index. In fact, Brown and Warner (1980) in their survey of event study methodologies observe that the use of an equally weighted market index is preferable to the use of a value weighted index in the context of market model based procedures. Chen, Roll, and Ross (1986) show that, to the extent that the market factor is priced in the capital markets, it is the *equally weighted* index rather than the *value weighted* index that is relevant. These empirical results are also consistent with the k-factor model presented in this paper.

II. Simulation Findings

Conway and Reinganum (1988) argue that a model such as the one described in the previous section is not empirically reasonable. A testable implication is that the second through kth eigenvalues are equal in magnitude. Conway and Reinganum (1988) and Trzcinka (1986) present empirical results based on New York Stock Exchange rate of return data that show that the second, third, and remaining sample eigenvalues differ significantly. Therefore, a model such as that described in the previous section cannot be responsible for security returns.

[10] The studies by King (1966), Brown and Weinstein (1985), and Chen, Copeland, and Mayers (1987) used factors generated by factor analysis rather than principal factor analysis. However, in the present example with equal idiosyncratic risk and a sufficient time series sample size, the first k principal factors will differ only by rotation from the k underlying factors. As is well known, the method of factor analysis yields factors unique only up to rotation. Most factor analysis algorithms use the first k principal factors as starting values. Therefore, in large time series samples the factors estimated by factor analysis should closely resemble the principal factors if the k-factor model is indeed correct.

This conclusion depends crucially on the little understood sampling properties of the sequence of estimated eigenvalues. To investigate this issue, a simulation experiment was conducted with the data assumed to be generated from the economy described in the previous section, with $k = 4$ equally important priced factors and a covariance matrix corresponding to observed data for the New York and American Stock Exchanges.

As expected, the first sample eigenvalue is large relative to the others as the number of securities grows large. However, the second, third, and remaining eigenvalues differ. Furthermore, this hypothetical example yields results quite close to those reported by Trzcinka (1986) for actual security returns.[11]

To study the sampling properties of the estimated eigenvalues in the context of the model of the previous section, eighty weeks of return data were generated for eighty securities. Returns were generated from an exact four-factor model. The factor loadings b_{ij}, $i = 1{,}80$, $j = 1{,}4$, were of a similar order of magnitude across securities and were each randomly chosen from a normal distribution with unit mean and a variance of 0.01. These factor loadings could be thought of as the factor loadings of eighty securities chosen at random from the set of all securities.

The returns for security i in period t were generated from

$$\tilde{R}_{it} - \mu_i = \sum_{j}^{4} b_{ij}\tilde{f}_{jt} + \tilde{e}_{it},$$

where $\tilde{f}_{jt} \sim \mathrm{NID}[0, 0.000158]$ and $\tilde{e}_{it} \sim \mathrm{NID}[0, 0.0045]$. The factor and idiosyncratic (weekly) standard deviations are chosen to conform to the Brown and Weinstein (1985) numbers, with an R^2 of 0.1235, also from Brown and Weinstein. Based on the analysis of the previous section, the derivative of the first eigenvalue with respect to the number of securities should equal 0.00063456 as the number of securities grows large, and the corresponding derivatives for the second through fourth eigenvalues should equal 0.000001585, while the derivatives for the remaining eigenvalues should equal zero. In strictly numerical terms, the second and smaller eigenvalues should be indistinguishable from zero.

Eigenvalues were computed on the basis of sample covariance matrices of returns on successively larger numbers of securities.[12] This experiment was repeated one hundred times, and the average value was computed for each eigenvalue and each number of securities. Figure 1 plots the largest ten eigenvalues as a function of the number of securities under consideration. We observe from this figure that the largest eigenvalue is significantly larger than the remaining eigenvalues and appears to increase without limit with the number of securities. The surprising result is that all remaining eigenvalues are larger than

[11] However, the example differs in the important respect that it assumes that the idiosyncratic terms have variance constant across securities. Trzcinka (1986) found that the idiosyncratic variance differed from security to security.

[12] The experimental design deliberately followed that of Trzcinka (1986) so as to preserve comparability with that work. As with Trzcinka (1986), the results reported in this section hold the number of observations constant as the number of securities increases. Additional work not reported here *increased* the number of observations with every security. This led to a small change in the numerical magnitude of the results reported in the paper but affected none of the qualitative results.

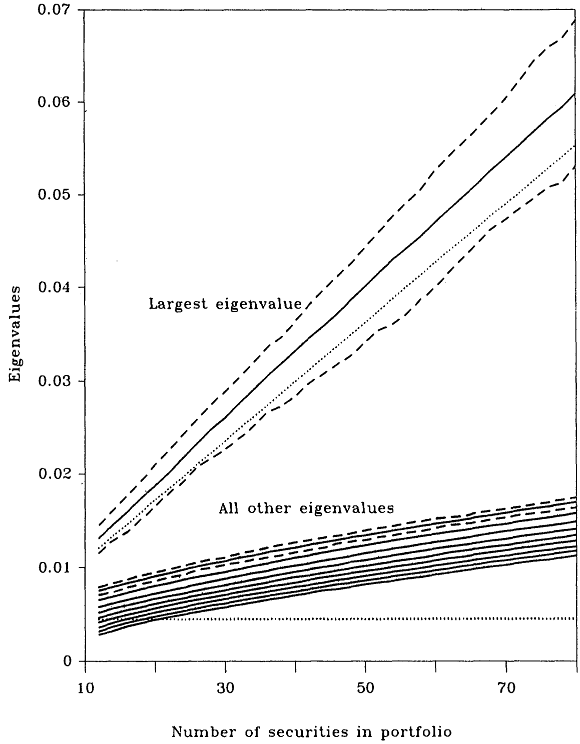

Figure 1. Ten largest eigenvalues as a function of portfolio size. In this figure, each solid line represents the sample mean of ranked eigenvalues estimated on the basis of 100 replications of the simulation experiment described in the text, for successively larger numbers of securities. The dashed lines represent the interquartile range of the empirical distribution of sample eigenvalues (for the first and second eigenvalues only). The dotted lines give the corresponding eigenvalues for the true or underlying covariance matrix of returns. Note that the second through tenth eigenvalues of the true covariance matrix are indistinguishable in this figure, corresponding to the numerical values reported for derivatives computed in Section II of the text.

expected. They differ significantly in value, and all, including eigenvalues smaller than the fourth, appear to increase with the number of securities. Thus, Conway and Reinganum's characterization of the empirical evidence may say more about the sampling properties of eigenvalues than about the underlying process generating security returns.

These results are reinforced by examining Panel A of Table I, which reports results from regressing each of the six largest eigenvalues on the number of securities. These results are remarkably similar to those reported by Trzcinka (1986) computed on the basis of observed stock return data and which are reported in Panel B of Table I. The slope coefficient for the first eigenvalue computed on the basis of the simulated data is approximately the same as the predicted value, 0.00063456. However, the slope coefficient for the actual data is 0.00048. While of a similar order of magnitude, this estimate is significantly smaller than the predicted value (with a t-value of -5.22). The slope coefficients

Table I

Regressing Eigenvalues Against the Number of Securities in the Portfolio (Nested Sequence of Securities)

The results in Panel A are the average value of coefficients in the regression:

$$e_{im} = \alpha_i + \beta_i \times m + \varepsilon_{im};$$
$$m = 12, 14, 16, \ldots, 78, 80, i = 1, \ldots, 6,$$

where i indexes eigenvalues, m indexes the number of securities upon which the ith eigenvalue was computed, and e_{im} represents the value of the ith eigenvalue computed on the basis of m securities, where the average is taken over 100 replications of the simulation experiment described in the text. The t-values are computed using the standard error of the sample mean coefficient, and the R^2 is the average value of the coefficient of multiple determination over these replications. The results in Panel B are taken directly from Trzcinka (1986).

A. Results for Simulated Four Factor Structure, 80 Securities			
Eigenvalue	Intercept[a]	Slope[a]	R^2
1	0.498 (24.4)	0.070 (57.2)	0.995
2	0.653 (84.0)	0.014 (93.4)	0.976
3	0.550 (94.7)	0.013 (118.6)	0.981
4	0.475 (103.9)	0.013 (154.7)	0.984
5	0.420 (112.2)	0.013 (175.9)	0.985
6	0.372 (104.8)	0.013 (188.4)	0.985
B. Results for Actual Data 1963–1983, 865 Securities			
Eigenvalue	Intercept[a]	Slope[a]	R^2
1	0.448 (2.8)	0.048 (16.2)	0.999
2	0.704 (21.3)	0.004 (6.3)	0.998
3	0.514 (11.7)	0.003 (35.7)	0.994
4	0.522 (11.1)	0.002 (22.5)	0.987
5	0.463 (10.8)	0.002 (24.5)	0.987
6	0.524 (10.3)	0.002 (20.9)	0.969

[a] All coefficients are multiplied by 100. t-Values are in parentheses.

of the remaining eigenvalues are all significantly larger than predicted. In particular, the slopes for eigenvalues five and six are significantly positive. These coefficients should be zero if a four-factor model is indeed responsible for the data, and the eigenvalues are measured without error. The slopes for eigenvalues one through twenty are all significant at the one percent level for the actual data (Trzcinka (1986)). In addition, it is difficult to observe a significant difference between eigenvalues five and six and eigenvalues two, three, and four when expressed as a percentage of the trace for successively larger portfolios of securities (Table II). The correspondence between Table II and the results reported by Trzcinka (1986) (given in the case of 100 securities on the last line of Table II) is quite remarkable.

It is important to note here that the simulation results reported in Figure 1

Table II

Six Largest Eigenvalues Expressed as a Ratio of the Trace of the Covariance Matrix, as the Number of Securities Increases

Simulated Data refers to the mean eigenvalues computed on the basis of 100 replications of the experiment described in the text. The numbers correspond to the numbers reported in Table I. The ratio of the trace is defined for the simulated data as the ratio of the mean eigenvalue to the trace of the covariance matrix from which the data were simulated to have come. Actual Data refers to results for 1963–1983 taken directly from Trzcinka (1986). In this case, the trace is taken of the sample covariance matrix.

Portfolio Size 20 (Simulated Data)						
	1	2	3	4	5	6
Eigenvalues	0.0188	0.0091	0.0081	0.0073	0.0067	0.0061
Ratio of Trace	0.1831	0.0886	0.0787	0.0707	0.0652	0.0599
Portfolio Size 40 (Simulated Data)						
	1	2	3	4	5	6
Eigenvalues	0.0332	0.0121	0.0111	0.0102	0.0096	0.0090
Ratio of Trace	0.1616	0.0591	0.0539	0.0497	0.0465	0.0437
Portfolio Size 60 (Simulated Data)						
	1	2	3	4	5	6
Eigenvalues	0.0470	0.0147	0.0136	0.0128	0.0120	0.0114
Ratio of Trace	0.1523	0.0477	0.0441	0.0414	0.0391	0.0368
Portfolio Size 100 (Actual Data)						
	1	2	3	4	5	6
Eigenvalues	0.0519	0.0109	0.0078	0.0075	0.0066	0.0065
Ratio of Trace	0.1922	0.0405	0.0289	0.0279	0.0244	0.0240

and Table I extend to consideration of portfolios comprising a maximum of 80 securities. As a practical matter, it is difficult to maintain numerical accuracy in the eigenvalue computations for portfolios of significantly higher dimension using standard mathematical software. Trzcinka (1986), with advanced numerical algorithms and significant computer resources, was able to extend the maximum dimension to 856 securities. As stated before, Trzcinka found results remarkably close to the results reported here. It is at least possible in theory to extend the analysis to much larger portfolios for which the investigator might be able to discern differences between the kth and k + 1st eigenvalues. Such differences, small as they are, do exist. However, such a possibility appears to lie outside the range of the available computer technology.

A possible conclusion one might draw from the Trzcinka findings is that, in addition to one large factor, there is a multiplicity of smaller pervasive factors responsible for security returns. This simulation experiment suggests that the small sample properties of eigenvalues may actually lead to a conclusion of "too many" factors. This multiplicity of factors as the number of securities increases has been observed by others using standard chi-square tests for the number of

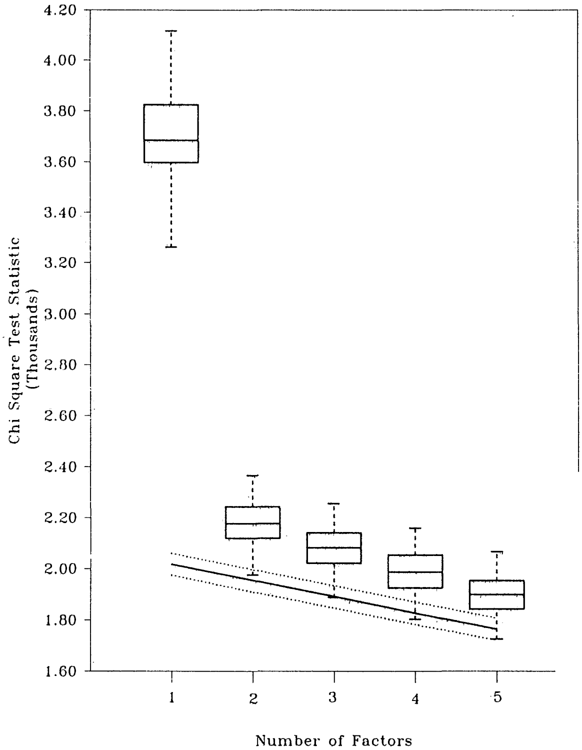

Figure 2. The sampling distribution of chi-square test statistics for the number of factors with 65 securities and 80 observations in 100 replications. These boxplots give the median (horizontal line), interquartile range (box), and 5 and 95 percentiles (whiskers) of the empirical distribution of the chi-square test statistic (with Bartlett correction) for the four-factor model described in the text. The sloped solid line gives the median of the theoretical distribution, and the dotted lines represent the interquartile range of that distribution.

factors.[13] It is interesting to discover whether these two sets of findings are associated.

To investigate this issue, standard chi-square tests for the number of factors were studied for the case of 65 securities. In Figure 2 we give boxplots of the empirical distribution of the test statistic. The interquartile range (given by the box in the figure) indicates that there is a large difference between the results for a test of one factor and tests of two or more factors. This is consistent with there being just one statistical factor. However, the interquartile ranges for *all* test statistics are greater than the theoretical values. If we were to take these results at face value, one might conclude that there are *more* than four factors responsible for the returns simulated for this example. While this is just another manifestation of the misspecification of the chi-square test applied to stock return data documented by Conway and Reinganum (1988) among others, it is

[13] See, for example, Dhrymes, Friend, and Gultekin (1984).

interesting that the same numerical example that leads the eigenvalue procedure to infer a multiplicity of small factors also has the chi-square procedure overestimate the number of factors.[14]

III. Conclusion

The verdict on the use of purely statistically based procedures to determine the number of factors in the context of actual stock market data appears to be unfavorable. For an example constructed such that each of k factors is priced and contributes equally to variance, the approximate factor solution suggests that only one factor is responsible for security returns. That one factor is the equally weighted market index.

This approximate factor solution is a unique, consistent estimator of an underlying factor structure that differs from the original factor structure by a particular nonsingular rotation. This so-called Helmert rotation has a number of interesting properties. While it is well known that the APT is invariant to such rotations, generally it is not understood that pricing tests depend on the rotation. The Helmert rotation has the property that it can increase the statistical significance of the first estimated factor price and reduce the statistical significance of the remaining factor price estimates. This suggests a CAPM bias in tests of the asset pricing relationships, a bias that is easily corrected by the reverse Helmert rotation.

There also appear to be small sample inference problems associated with the sequence of estimated eigenvalues that in certain circumstances might lead the empirical investigator to the false conclusion that there are *too many* factors. These results are consistent with problems others have found interpreting chi-square tests for the number of factors.

It is important to note that this result is not due to any kind of analytical error in Chamberlain and Rothschild (1983), which first suggested the analysis of the approximate factor solution in this context. It is simply that the use of such procedures can lead to ambiguous conclusions where empirical investigators apply them to the study of finite (and in many cases small) numbers of securities. To put this point another way, suppose one started with a prior conviction that there was one major factor responsible for security returns, modified perhaps by several minor factors. Then one might find comfort in these results. The empirical evidence based on the analysis of eigenvalues is consistent with such a view. Does this mean that, when all is said and done, the CAPM remains an appropriate specification of equilibrium in the capital markets? Unfortunately, this paper shows that the empirical evidence is *also* consistent with a k-factor model where each factor is equally important and is priced in an APT context. It would be wrong, of course, to assert that, simply because the statistical procedures lack

[14] We also considered a case where the data were simulated to correspond to the distribution found by Huberman and Kandel (1985). This case corresponded to an example where there were three equally important (but correlated) factors. Consistent with the findings of Conway and Reinganum (1988), the first two eigenvalues could be distinguished from the remaining eigenvalues. Furthermore, consistent with the other example, the first eigenvalue was significantly larger than the second and remaining eigenvalues.

sufficient power to identify more than one factor, only one factor is priced. Such an error would mirror in a curious way arguments made by others that, because chi-square tests appear to identify a multiplicity of factors, each of those "factors" must be priced. (Roll and Ross (1984) discuss such arguments.) Thus, mechanical application of purely statistical approaches to determining the number of pervasive factors in equity returns may lead to false inferences. Unfortunately, it seems that economic analysis and intuition are essential ingredients to this process.[15]

[15] See, for example, the work of Chen, Roll, and Ross (1986).

REFERENCES

Brown, Stephen J. and Jerold Warner, 1980, Measuring security price performance, *Journal of Financial Economics* 8, 205–258.

——— and Jerold Warner, 1985, Using daily stock returns: The case of event studies, *Journal of Financial Economics* 14, 3–31.

——— and Mark I. Weinstein, 1983, A new approach to testing asset pricing models: The bilinear paradigm, *Journal of Fiance* 38, 711–743.

——— and Mark I. Weinstein, 1985, Derived factors in event studies, *Journal of Financial Economics* 14, 491–495.

Chamberlain, Gary and Michael Rothschild, 1983, Arbitrage and mean variance analysis on large markets, *Econometrica* 51, 1281–1301.

Chen, Nai-fu, 1983, Some empirical tests of the theory of arbitrage pricing, *Journal of Finance* 38, 1393–1414.

———, Richard Roll, and Stephen Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

———, Thomas Copeland, and David Mayers, 1987, A comparison of single and multifactor portfolio performance methodologies, *Journal of Financial and Quantitative Analysis* 22, 401–417.

Connor, Gregory and Robert Korajczyk, 1988, Risk and return in an equilibrium APT: Application of a new test methodology, *Journal of Financial Economics* 21, 255–289.

Conway, Dolores and Marc Reinganum, 1988, Stable factors in security returns: Identification using cross validation, *Journal of Business and Economic Statistics* 6, 1–15.

——— and Marc Reinganum, 1988, Reply, *Journal of Business and Economic Statistics* 6, 24–28.

Dhrymes, Phoebus, Irwin Friend, and N. Bulent Gultekin, 1984, A critical examination of the empirical evidence of the arbitrage pricing theory, *Journal of Finance* 39, 323–346.

Huberman, Gur and Shmuel Kandel, 1985, A size based asset pricing model, 1985, Unpublished working paper, University of Chicago.

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

King, Benjamin, 1966, Market and industry factors in stock price behavior, *Journal of Business* 39, 139–190.

Linn, Scott and E. Chang, 1985, The empirical implications of approximate factor structure in large asset markets, Unpublished working paper, University of Iowa.

Luedecke, B., 1984, An empirical investigation into arbitrage and approximate k-factor structure in large asset markets, Unpublished Ph.D. dissertation, University of Wisconsin.

Markowitz, Harry, 1959, *Portfolio Selection: Efficient Diversification of Investments* (John Wiley & Sons, New York).

Press, S. James, 1972, *Applied Multivariate Analysis* (Holt, Rinehart and Winston, New York).

Roll, Richard, 1977, A critique of the asset pricing theory's tests, *Journal of Financial Economics* 4, 129–176.

———, 1985, A note on the geometry of Shanken's CSR T^2 test for mean/variance efficiency, *Journal of Financial Economics* 14, 349–357.

——— and Stephen Ross, 1980, An empirical investigation of the Arbitrage Pricing Theory, *Journal of Finance* 35, 1073–1102.

——— and Stephen Ross, 1984, A critical reexamination of the empirical evidence on the Arbitrage Pricing Theory: A reply, *Journal of Finance* 39, 347–350.

Ross, Stephen, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 3, 343–362.

Shanken, Jay, 1982, The arbitrage pricing theory: Is it testable? *Journal of Finance* 37, 1129–1140.

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

Sharpe, William, 1963, A simplified model for portfolio selection, *Management Science* 9, 277–293.

Tiemann, Jonathan, 1988, Exact arbitrage pricing and the minimum-variance frontier, *Journal of Finance* 43, 327–338.

Trzcinka, Charles, 1986, On the number of factors in the arbitrage pricing model, *Journal of Finance* 41, 347–368.