ELSEVIER

Journal of Empirical Finance 2 (1995) 71–93

Journal of
EMPIRICAL
FINANCE

Small sample rank tests with applications to asset pricing

Guofu Zhou [*,1]

Washington University, St. Louis, MO 63130, USA

Final version received October 1994

Abstract

This paper proposes small sample tests for rank restrictions that arise in many asset pricing models, economic fields and others, complementing the usual asymptotic theory which can be unreliable. Using monthly portfolio returns grouped by industry and using two sets of instrumental variables, we cannot reject a one-factor model for the industry returns.

Keywords: Rank test; Small sample test; Asset pricing; Factor model

JEL classification: G12; G11; C11; C31

1. Introduction

The regression coefficient matrix is often assumed to be of full rank in the classical multivariate regression model. If the rank is not full, then we have the so called reduced rank regression model. To test for the regression coefficient matrix to have a given rank, Anderson (1951) derived in his seminal contribution, maximum likelihood estimators and the associated asymptotic test. Subsequently, Jöreskog and Goldberger (1975), Izenman (1975), Tso (1981), Davies and Tso

* *Correspondence to*: Guofu Zhou, John M. Olin School of Business, Washington University, St. Louis, MO 63130. USA.

[1] I am grateful to Phil Dybvig, Wayne Ferson, Ronald Gallant, John Geweke, Campbell Harvey, Bill McDonald, Sunil Panikkath, Peter Robinson, Jim Steeley, Jack Strauss, Arnold Zellner, Yao-Ting Zhang, and especially to an Associate Editor, two anonymous referees and Franz Palm (the Editor) for many helpful comments that substantially improved the paper. Financial support from the Fossett Foundation is also gratefully acknowledged.

SSDI 0927-5398(94)00011-5

(1982), among others, have developed further procedures for estimation and tests in the reduced rank regression model. The recent work of Velu (1992) has generalized this analysis to models with two sets of regressors where both have reduced rank restrictions. The statistical inference in all of these studies rely on the asymptotic chi-squared distribution of the likelihood ratio test. However, to the author's knowledge, there have not been any studies about the appropriateness of the asymptotic approximation in small sample size. Indeed, as demonstrated in this paper, the asymptotic chi-squared test can be unreliable. In some cases, for example, the asymptotic P-value is less than 0.001 whereas the exact P-value is greater than 0.10.

Although the reduced rank models were first motivated as an interesting statistical problem, they have wide applications in both economics and physical sciences. In macroeconomics, for example, Zellner (1970) and Sargent and Sims (1977) used the reduced rank models to study latent or unobservable factors and indices. Chen (1981) illustrated the application in physical sciences. In finance, there have been ample applications which may be partly due to the quantity and quality of data available. Hansen and Hodrick (1983) and Gibbons and Ferson (1985) used the reduced rank models to examine time-varying factor risk premiums. Subsequently, Campbell (1987), Ferson and Harvey (1991), among others, used similar models to study stock returns, forward currency premiums, international equity returns and capital integration. Stambaugh (1988) developed a latent variable model to test the Cox, Ingersoll and Ross' (1985) theory of the term structure of interest rates.

This paper proposes a Wald test whose exact distribution in small samples can be computed by using a simple numerical approach. The test is in fact equivalent to the usual likelihood ratio test, and so it enjoys the usual interpretations. Because many asset pricing models imply the same rank restrictions, wide applications of the test are possible. As an example, we investigate the number of 'priced' latent factors in the US equity market by using monthly portfolio returns grouped by industry. Based on two sets of instrumental variables, we cannot reject the hypothesis that a one-factor model explains the industry returns reasonably well.

In comparison with the commonly used general method of moments (GMM) (Hansen (1982)), our approach has at least two important advantages. First, the inference is exact. There is no concern about whether the asymptotic distributions are appropriate or not. Second, the sample size can be small which allow us to examine an asset pricing model over a short time horizon. (The GMM approach cannot be used in our applications to five-year subperiods because Hansen's optimal weighting matrix is singular even with doubling of the sample size.) On the other hand, the GMM procedure allows the disturbances to have conditionally time-varying distributions which is sensible in a long time period. But our approach allows time-varying parameters across subperiods and it should capture some of the time-varying effects. So, our approach can be a good alternative to the GMM procedure for testing many asset pricing models.

The paper is organized as follows. In Section 2, we show that many asset pricing models imply rank restrictions on regression coefficients. In Section 3, we propose the Wald test, provide a simple method for the computation of its exact P-value as well analytical bounds, and discuss its relation with the likelihood ratio test. In Section 4, we investigate the number of latent factors in the US equity market. Section 5 concludes the paper.

2. Rank restrictions and asset pricing constraints

In this section, we introduce the rank restriction in general and then give examples showing that a rank restriction is an implication of many asset pricing models. Therefore, the rank test developed in this paper may be used to examine the validity of a wide class of asset pricing models.

Suppose N financial variables $Y_1,\ldots,Y_N$ are related to L explaining variables $X_1,\ldots,X_L$ by the traditional multivariate regression model:

$$Y_{it} = \theta_{i1}X_{1t} + \cdots + \theta_{iL}X_{Lt} + \epsilon_{it},\ i = 1,\ldots,N,\ t = 1,\ldots,T, \tag{1}$$

where T is number of periods (sample size), and ϵ_{it} is the model disturbances or forecasting errors. Because we often try to use a small number of variables to explain Y (say, using a few economic variable to explain the hundreds security returns), it is reasonable to assume that the number of explaining variables is less than or equal to the number of variables to be explained, i.e., $L \leqslant N$. This implies that the $L \times N$ parameter matrix formed by the regression coefficients of model (1),

$$\Theta \equiv \begin{pmatrix} \theta_{11} & \theta_{21} & \cdots & \theta_{N1} \\ \theta_{12} & \theta_{22} & \cdots & \theta_{N2} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{1L} & \theta_{2L} & \cdots & \theta_{NL} \end{pmatrix},$$

must have a rank of at most L. As a result, the columns must lie in a space of dimension no greater than L. As shown later in this section, many asset pricing models predict that the columns lie in a space of dimension K where K is some prespecified number less than L. This translates into the following rank restriction on the regression coefficient matrix:

$$H_0\colon\ \text{rank}(\Theta) = K. \tag{2}$$

The procedures for testing the validity of this rank restriction are developed in the next section. Now we examine the relationships between the rank restriction and the implications of many asset pricing models.

2.1. Intertemporal factor pricing models

Following Gibbons and Ferson (1985), a general intertemporal K-factor asset pricing model has the following form: [2]

$$E(r_{it}|\mathbf{Z}_{t-1}) = \lambda_0(\mathbf{Z}_{t-1}) + \beta_{i1}\lambda_1(\mathbf{Z}_{t-1}) + \cdots + \beta_{iK}\lambda_K(\mathbf{Z}_{t-1}), \tag{3}$$

where r_{it} is the return on asset i between period $t-1$ and t, $\lambda_j(\mathbf{Z}_{t-1})$ the market-wide expected risk premium on the j-th factor (of the K latent factors), $\mathbf{Z}_{t-1} = (Z_{1,t-1}, \ldots, Z_{L,t-1})'$ the market-wide information available at t, and $\beta_{i1}, \ldots, \beta_{iK}$ the conditional betas for asset i. The model allows $\lambda_j(\mathbf{Z}_{t-1})$'s to vary over time but requires β's to be constant. In terms of excess returns, (3) can be written:

$$E(R_{it}|\mathbf{Z}_{t-1}) = b_{i1}\lambda_1(\mathbf{Z}_{t-1}) + \cdots + b_{iK}\lambda_K(\mathbf{Z}_{t-1}), \tag{4}$$

where $R_{it} = r_{it} - r_{0t}$ is the return on the i-th asset in excess of the return on the 0-th asset (chosen arbitrarily), and $b_{ij} = \beta_{ij} - \beta_{0j}$ is the 'excess' conditional beta. To test (4), we assume as in most studies the following returns-generating process:

$$R_{it} = \theta_{i1}Z_{1,t-1} + \cdots + \theta_{iL}Z_{L,t-1} + \epsilon_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T. \tag{5}$$

It can be shown that the K-factor model restriction (4) is valid if and only if the $L \times N$ regression coefficient matrix Θ of (5) has rank K. Therefore, a test of the K-factor pricing theory is a test of the rank restriction.

2.2. One-period factor asset pricing models

In one-period asset pricing models, rather than (5), the returns or excess returns (in excess of a risk-free rate or the return on some zero-beta asset) are often assumed to have a factor structure:

$$R_{it} = \alpha_i + b_{i1}f_{1t} + \cdots + b_{iK}f_{Kt} + \epsilon_{it},\ i = 1, \ldots, N, \tag{6}$$

where $f_{1t}, \ldots, f_{Kt}$ are the factors or the returns on some benchmark portfolios. When there exists a riskless asset, an asset pricing model which states that $E(\mathbf{R}) = \mathbf{B}\gamma$ for some factor premiums γ, implies (see, e.g., Shanken, 1985) the following restriction on the parameters of the linear regression (6):

$$\alpha = \mathbf{B}\lambda, \tag{7}$$

where α an $N \times 1$ vector of the alphas; $\mathbf{B}$ an $N \times K$ matrix of the bs; and $\lambda = \gamma - E(\mathbf{f})$ the risk premiums. Let Θ be the regression coefficients of (6), then (7) implies that Θ is not of full rank. Therefore, a rejection of Θ being of rank lower than or equal to K is a rejection of the asset pricing model.

[2] The notation in this and the following two subsections is used to follow the convention in finance. It is independent of the notation in the rest of the paper.

If a riskless asset does not exist, the asset pricing restriction has the form:

$$\alpha = \lambda_0(\mathbf{1}_N - \mathbf{B}\mathbf{1}_K), \tag{8}$$

where λ_0 is the return on the zero-beta asset, $\mathbf{1}_N$ and $\mathbf{1}_K$ are vectors of 1s. Consider the new regression:

$$R_{it} - (f_{1t} + \cdots + f_{Kt})/K = \alpha_i + (b_{i1} - 1/K)f_{1t} + \cdots + (b_{iK} - 1/K)f_{Kt} + \epsilon_{it}. \tag{9}$$

(8) implies that the regression coefficients of (9) must have a rank lower than or equal to K. Hence, a rank test can again be used to test the pricing restrictions.

A rejection of the rank restriction is a rejection of the model, but a failure to reject does not necessarily give rise to support for the model because the rank restriction is only a necessary condition. Gibbons (1982) was the first to develop an asymptotic test to test (8) directly without resorting to the rank test. Overcoming problems in using the asymptotic test, Shanken (1985, 1986) derived a small sample bound test. In the special case where $K = 1$ and the factor is the market portfolio, (8) reduces to the well known Black's (1972) zero-beta capital asset pricing model (CAPM) for which Zhou (1991) obtained an exact eigenvalue test. However, in the case where $K > 1$, the eigenvalue test is no longer applicable, but the rank test still applies.

The factors have so far been assumed to be observable, but they are not so in Ross' (1976) arbitrage pricing theory (APT). Nevertheless, we can still develop a rank test for the APT restrictions:

$$\alpha_i = \lambda_0 + b_{i1}\lambda_1 + \cdots + b_{iK}\lambda_K, \; i = 1, \ldots, N. \tag{10}$$

Motivated by Gibbons and Ferson (1985), we assume, without loss of generality, that the first $K \times K$ submatrix of $\mathbf{B}$, $\mathbf{B}_1$, is nonsingular. By (6), we obtain a regression of $\mathbf{R}_{2t}$ on $\mathbf{R}_{1t}$:

$$\mathbf{R}_{2t} = \left(\alpha_2 - \mathbf{B}_2\mathbf{B}_1^{-1}\alpha_1\right) + \mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{R}_{1t} + \mathbf{u}_t, \tag{11}$$

where $\mathbf{u}_t = -\mathbf{B}_2\mathbf{B}_1^{-1}\epsilon_{1t} + \epsilon_{2t}$. Partitioning (10) into $\alpha_1 = \lambda_0\mathbf{1}_K + \mathbf{B}_1\lambda$ and $\alpha_2 = \lambda_0\mathbf{1}_{N-K} + \mathbf{B}_2\lambda$, we get:

$$\alpha_2 - \mathbf{B}2\mathbf{B}_1^{-1}\alpha_1 = \lambda_0\left(\mathbf{1}_{N-K} - \mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{1}_K\right). \tag{12}$$

This is similar to (8), and hence a rank test can be used to test the APT.

2.3. Factor reduction, dimensionality reduction and multiple hedging

In contrast with (6), a slightly more general factor model for the asset returns is:

$$R_{it} = \theta_{i1}f_{1t} + \cdots + \theta_{iL}f_{Lt} + \epsilon_{it}, \; i = 1, \ldots, N, \tag{13}$$

where $f_{1t},\ldots,f_{Lt}$ are the factors (f_{1t} may be a constant). If Θ has rank K, there must exist an $L\times K$ matrix $\mathbf{A}$ and a $K\times N$ matrix $\mathbf{B}$ such that $\Theta=\mathbf{AB}$. Therefore, (13) implies that:

$$E(R_i)=b_{i1}E(f_1^*)+\cdots+b_{iK}E(f_K^*),\ i=1,\ldots,N, \tag{14}$$

where $(f_1^*,\ldots,f_K^*)$ are K 'new' factors that are linear combinations of the 'old' ones, $(f_1^*,\ldots,f_K^*)'=(f_1,\ldots,f_L)'\mathbf{A}$. Because any non-singular linear transformation, $\mathbf{AC}$ and $\mathbf{B}^{-1}\mathbf{B}$, gives rise to the same Θ, $\mathbf{A}$ and $\mathbf{B}$ cannot be uniquely identified; but they are determined up to that linear transformation. If $\mathbf{A}$ is normalized so that $(f_1^*,\ldots,f_K^*)$ are portfolios of the "old" factors $f_1,\ldots,f_L$, $\mathbf{B}$ is uniquely determined and is the matrix of the risk-premiums associated with the "new" factors. Then, (14) states that the expected returns on the assets are rewarded by only K priced factors (fewer than L). So, a test for the rank of Θ can help to reduce the number of factors that are necessary to describe the returns.

Although the rank test is motivated by a desire to test asset pricing theories, it also serves as a useful tool for hedging and econometric data modelling. The rank test directly answers the question whether there exist K portfolios from L available factors (say, the price changes of L futures contracts) so that a hedge using the K portfolios can do as well as a hedge using all the L factors. In data modelling, the greater the number of regressors, the better the model fits the data, but the worse the sampling behavior of the parameter estimates. So, for a given level of goodness of fit, we would like the number of explaining variables to be at a minimum and the rank test determines this minimum number K.

3. The econometric methodology

3.1. A wald test

To motivate, consider the unconstrained estimator of the model parameters. For notational brevity, we in what follows use the matrix form of model (1):

$$\mathbf{Y}=\mathbf{X}\Theta+\mathbf{E}, \tag{15}$$

where $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{E}$ are $T\times N$, $T\times L$, and $T\times N$ matrices respectively. As usual, we assume $\mathbf{X}'\mathbf{X}$ is nonsingular, and the disturbances, each row of $\mathbf{E}$, are independent and identically distributed multivariate normal with zero mean and nonsingular covariance matrix Σ. The thetas in the i-th equation are easily estimated from an OLS regression and the results can be written in matrix form:

$$\hat{\Theta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},\ L\times N. \tag{16}$$

The estimator of Σ is given by the cross products of residuals divided by T:

$$\hat{\Sigma}=\frac{1}{T}(\mathbf{Y}-\mathbf{X}\hat{\Theta})'(\mathbf{Y}-\mathbf{X}\hat{\Theta}),\ N\times N. \tag{17}$$

Notice that $\hat{\Theta}$ and $\hat{\Sigma}$ are also the maximum likelihood estimator (MLE) under full rank model, and are both consistent and efficient.

The vector formed by the rows of $\hat{\Theta}$ has a covariance matrix $(\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma^{-1}$ which can be estimated by $(\mathbf{X}'\mathbf{X})^{-1} \otimes \hat{\Sigma}^{-1}$. Standardizing $\hat{\Theta}$, we get a matrix form of the usual Wald test for the hypothesis that Θ is zero: [3]

$$\hat{\mathbf{W}} \equiv (\mathbf{X}'\mathbf{X}/T)^{1/2} \hat{\Theta} \hat{\Sigma}^{-1} \hat{\Theta}' (\mathbf{X}'\mathbf{X}/T)^{1/2}, \; L \times L \tag{18}$$

If Θ is zero, this matrix should be close to a zero matrix. However, here we are interested in the rank of Θ. Nevertheless, the same intuition is helpful. If Θ is of rank K, its estimator $\hat{\Theta}$ should be close to a matrix of rank K, and so should $\hat{\mathbf{W}}$. In this case, we would expect that the $(L-K)$ smallest eigenvalues of $\hat{\mathbf{W}}$ are small. Formally, let $\xi_1 \geq \xi_2 \geq \cdots \geq \xi_L \geq 0$ be the eigenvalues of $\hat{\mathbf{W}}$, and

$$W_K \equiv \prod_{j=K+1}^{L} (1 + \xi_j) \tag{19}$$

be the test statistic for the null hypothesis that the rank of Θ is K:

$$H_0: \quad \text{rank}(\Theta) = K. \tag{2}'$$

As an increasing function of the variables, this statistic summarizes the information on the deviations of the eigenvalues from zero. If W_K is large, it must be the case that at least one of the eigenvalues is large, implying that $\hat{\mathbf{W}}$ is likely to have a rank greater than K. Hence, we would reject the null hypothesis that Θ is of rank K. To assess the error of making a false rejection, we need to find W_K's probability distribution under the null hypothesis. This will be the focus of the paper, although our approach can yield similarly the test's distribution under the alternative hypothesis that the rank of Θ is greater than K.

To obtain the distribution of W_K under the null, we need to know the distribution of the eigenvalues. Let $\Delta = \text{diag}(\delta_1, \ldots, \delta_L)$ be the roots of:

$$|(\mathbf{X}'\mathbf{X}/T)^{1/2} \Theta \Sigma^{-1} \Theta' (\mathbf{X}'\mathbf{X}/T)^{1/2} - \delta \mathbf{I}_L / T| = 0, \tag{20}$$

where $\delta_1 \geq \cdots \geq \delta_L \geq 0$, and denote by $W_L(N, \mathbf{I}_L, \Delta)$ and $W_L(T-N, \mathbf{I}_L)$ the density functions of a non-central and a central Wishart distributions, we have then:

Theorem 1. If $N \geq L$, $T \geq N + L$, the joint density of $\xi_1, \ldots, \xi_L$ has exactly the same form as the joint density of the eigenvalues of $\mathbf{U}\mathbf{V}^{-1}$, where $\mathbf{U} \sim W_L(N, \mathbf{I}_L, \Delta)$ and $\mathbf{V} \sim W_L(T-N, \mathbf{I}_L)$.

Proof: See Appendix A.

Theorem 1 provides in principle the distribution of the Wald test for rank hypotheses. Interestingly, it appears exactly like Theorem 10.4.2 of Muirhead

[3] In an actual computation of the Wald matrix $\hat{\mathbf{W}}$, it may be more convenient if $(X'X/T)^{1/2}$ is replaced by the L matrix of the standard Cholesky LU-decomposition for the matrix $(X'X/T)$.

(1982), showing that the distribution of the rank test is determined by the joint density of the eigenvalues of $\mathbf{UV}^{-1}$, which also determines the Wald test for linear hypotheses. However, in comparison with the test for linear hypotheses, the rank test is computed differently from the data and its distribution is determined by the eigenvalues in different functional form. Hence, Theorem 1 should be regarded as different from Muirhead's Theorem 10.4.2 in the sense that the former applies to the rank test, while the latter is developed with the purpose of testing linear hypotheses.

Based on Theorem 1, a three step approach may be taken to draw W_K from its exact distribution at a given value of $(\delta_1, \ldots, \delta_L)$. First, since the deltas are given, random samples of $\mathbf{U}$ and $\mathbf{V}$ can be drawn from $W_L(T-N, \mathbf{I}_L)$ and $W_L(N, \mathbf{I}_L, \Delta)$ (see Appendix B for a detailed description). Second, after each draw, the eigenvalues of $\mathbf{UV}^{-1}$ can be computed by using standard eigenvalue routines. Third, the test statistic W_K is then evaluated based on the eigenvalues from step 2. This W_K will be one random draw from its exact distribution. If we repeat this process 10,000 times, for example, we will obtain 10,000 random draws of W_K, and hence the percentage of times that W_K is less than a given number x is readily computed. This percentage is a numerical approximation to $\text{Prob}(W_K < x)$ at the given $(\delta_1, \ldots, \delta_L)$. In other words, the exact cumulative distribution function of the Wald test statistic W_K is easily computed by using the Monte Carlo approach just described. It is worth emphasizing that this approach is different from the usual Monte Carlo simulation approach. To use the latter, the model parameters have to be prespecified, the residuals have to be generated, and then hundreds and thousands of parameter estimations have to be performed. In contrast, our approach is done by drawing samples from two Wishart distributions, providing a simple, generic and very practical way to obtain the exact distribution of the Wald test W_K.

Alternatively, the above approach can be viewed as the standard Monte Carlo integration approach applied to evaluate the integral:

$$\mathrm{P}(W_K < x) = \int \cdots \int g(\xi_{K+1}, \ldots, \xi_L) f(\xi_{K+1}, \ldots, \xi_L) \, d\xi_{K+1} \cdots d\xi_L, \tag{21}$$

where $\mathrm{f}(\xi_{K+1}, \ldots, \xi_L)$ is the density function of $\xi_{K+1}, \ldots, \xi_L$, and $\mathrm{g}(\xi_{K+1}, \ldots, \xi_L)$ is an indicator function such that $\mathrm{g}(\xi_{K+1}, \ldots, \xi_L) = 1$ if $W_K < x$; and zero otherwise. If $f(\xi_{K+1}, \ldots, \xi_L)$ is easily computed, the integral can be computed by any standard numerical method, but Monte Carlo integration is preferred because a grid method will have difficulties for large dimensional problems. Now it is in fact very difficult to compute $f(\xi_{K+1}, \ldots, \xi_L)$ because it is an integral of matrix zonal polynomials (Muirhead, 1982) and it is well known that the computation of the matrix zonal polynomials is extremely complex (almost intractable when the dimensionality is greater than two). As a result, Monte Carlo integration becomes the only choice to compute (21). The integral is approximated by the sample mean

of $g(\xi_{K+1},\ldots,\xi_L)$ evaluated at draws of $\xi_{K+1},\ldots,\xi_L$ from its distribution $f(\xi_{K+1},\ldots,\xi_L)$. The draws are accomplished easily, and are exactly done by using the three step approach. Clearly, the sample mean of $g(\xi_{K+1},\ldots,\xi_L)$ is exactly equal to the previous percentage.

One advantage of viewing the numerical approach as Monte Carlo integration is that standard techniques about the latter (see, e.g, Geweke, 1989) may be used to assess the numerical error of the former. It follows that the rate of convergence of the numerical approximation to its exact value is independent of T, N, L and K, and is only a function of the number of the Monte Carlo draws. The greater the number of the draws, the more accurate the numerical approximation. Moreover, as is well known, 10,000 Monte Carlo draws generally deliver results that are accurate up to 2 to 3 decimal points.

In practice, the P-value is of interest for making inference. Under the null hypothesis that the rank restriction holds, $\delta_{K+1} = \ldots = \delta_L = 0$. However, we do not know $\delta_1,\ldots,\delta_K$, which are nuisance parameters on which the P-value depends. Nevertheless, the P-value can be computed at the maximum likelihood estimator (MLE) of $\delta_1,\ldots,\delta_K$ obtained from (34) with Θ and Σ being replaced by their constrained ML estimators (Section 3.3). The effect of using estimated deltas can be easily examined because the P-value is an increasing function of each of the deltas (by Theorem 1 and Perlman and Oklin (1980)). In other words, if the P-value is evaluated at a value of the deltas lower than the true one, we will underestimate the true probability; and if evaluated at a value higher than the true deltas, we will overestimate the true probability. Nevertheless, numerical experiments show that there are no substantial changes in the P-value when small perturbations in the deltas are allowed. For example, as shown later in our empirical applications, the first P_W entry, 0.1688, of Table 3 is the P-value evaluated at the MLE of δ_1 ($\tilde{\delta}_1 = 34.6367$). If re-evaluated at $\delta_1 \pm 20$ (sizable perturbations of δ_1), the P-value becomes 0.1651 or 0.1693, changed only slightly from 0.1688. For this insensitivity, it is worth mentioning that there is a theoretical reason that the asymptotic distribution of W_K does not depend on the nuisance parameters.

Because the P-value is computed at the ML estimates, rather than at the true but unknown value of $(\delta_1,\ldots,\delta_K)$, this introduces some uncertainty about the true P-value. However, as the previous example shows, the uncertainty can be easily quantified by utilizing the monotone dependence of the P-value on the deltas, and the uncertainty may not matter that much due to the insensitivity of the P-value to small perturbations in the deltas. Nevertheless, it is of interest to run a Monte Carlo experiment to see how the P-value computed from the ML estimates performs. In other words, if we fix the size of the test at the 5% and 10% level, we want to see whether the empirical rejection rates are close to 5% and 10% in a simulation where hundreds sets of the data are generated. To generate the data sets, the parameters Θ and Σ have to pre-specified. For simplicity, they are set equal to the ML estimates based on actual returns data (see Section 4) under the

rank one restriction. Given the parameter specifications, it is straightforward to generate hundreds sets of the residuals and the returns. Notice that the P-value has to be evaluated by Monte Carlo integration for each of the data set because the ML estimates of the deltas are different with different data sets. To make the computation manageable, we use only 2,000 simulated data sets, and use only 1,000 Monte Carlo draws for evaluating the P-value at each data set. When the sample size is $T = 60$ (a 5-year period for monthly data), the empirical rejection rates are 4.2% and 7.6%, remarkably close the desired 5% and 10% level. When we double the sample size to $T = 120$, the empirical rejection rates are 4.5% and 8.5%, becoming much closer to the 5% and 10% level. To summarize, the simulation provides evidence for the reliability of the P-value computed from the MLE of the deltas, although it is performed only in a couple of special cases due to the computational time required. Once again, in a more general situation where the simulation result is not directly applicable, the monotone dependence of the P-value on the deltas can help to assess the uncertainty caused by computing it at the ML estimates of the deltas.

3.2. *Analytical bounds and asymptotic distributions*

The monotonicity of the P-value suggests that if the limit of $\text{Prob}(W_K > x)$ exists as all the δ_i's go to infinity, it must be an upper bound on the P-value. Furthermore, this bound is *optimal* in the sense that, among all distributions that do not depend on the delta parameters, this limit is the tightest upper bound possible. Fortunately, we can obtain this tightest upper bound explicitly:

Theorem 2. Under the null hypothesis that the rank restriction (2) is true, if $N \geqslant L$, $T \geqslant N + L$, the upper bound on the P-value is:

$$\text{P}(W_K > x) \leqslant \text{P}(U > x), \quad \forall x > 0, \tag{22}$$

where $U = |\mathbf{W}_2| / |\mathbf{W}_1 + \mathbf{W}_2|$, $\mathbf{W}_1$ and $\mathbf{W}_2$ are independent, $\mathbf{W}_1 \sim W_{L-K}(N - K, \mathbf{I}_{L-K})$ and $\mathbf{W}_2 \sim W_{L-K}(T - N, \mathbf{I}_{L-K})$.

Proof. The proof following from our Theorem 1 and Schott (1984) by replacing $Q_1 = W$, $Q_1^* = U$, $s = K + 1$, $m = L$, $h = N$, and $e = T - N$ in his Theorem 3. **Q.E.D.**

The upper bound does not contain non-central delta parameters and can be easily computed, similar to the P-value, by Monte Carlo integration. If an upper bound of 4% is found, for example, the true significance level must be less than or equal to 4%, so we can reject the null at the usual 5% level.

By using again the monotonicity of the P-value, we know that $\text{Prob}(W_K > x)$ evaluated at zero deltas must be the optimal lower bound of the P-value. This lower bound is useful for accepting the null hypothesis. If a lower bound of 6% is found, for example, the true significance level must be at least 6%, then we cannot reject the hypothesis at the usual 5% level.

Explicit analytical expressions for both the optimal upper and lower bounds are

available in some special cases. They offer insights into the general properties of the test and serve as tools to verify solutions obtained by Monte Carlo integration. The first special case is when $K = L - 1$. The optimal upper bound is simply determined by an F-distribution (by Theorem 2 and Anderson, 1984, p. 305):

$$\frac{1-U}{U} \cdot \frac{T-N}{N-K} \sim F_{N-K,T-N}. \tag{23}$$

The lower bound is determined by the distribution of the smallest eigenvalue, ξ_L. Based on Nanda (1947), we have for $L = 2$,

$$\begin{aligned} \mathrm{P}(\xi_2 \geqslant x) = k_2\Big(&2B_{(1-x)}(2n_2+2,\ 2n_1+2) \\ &-x^{n_1+1}(1-x)^{n_2+1}B_{(1-x)}(n_2+2,\ n_1+1)\Big), \end{aligned} \tag{24}$$

and for $L = 3$,

$$\begin{aligned} &\mathrm{P}(\xi_3 \geqslant x) \\ &= k_3\Big(2B_{(1-x)}(2n_2+4,\ 2n_1+2)\cdot B_{(1-x)}(n_2+1,\ n_1+1) \\ &-2B_{(1-x)}(n_2+2,\ n_1+1)B_{(1-x)}(2n_2+3,\ 2n_1+2) \\ &-x^{n_1+1}(1-x)^{n_2+2}l_2(x)\Big), \end{aligned} \tag{25}$$

where $n_1 = (N-L-1)/2$, $n_2 = (T-N-L-1)/2$, $B_x(a,\ b)$ is the incomplete beta function: $B_x(a,\ b) \equiv \int_0^x t^{a-1}(1-t)^{b-1}\,dt$, and k_3 is a constant:

$$\begin{aligned} k_3 = &\frac{2\pi^{L/2}}{n_1+n_2+L}\prod_{i=1}^{L}\Gamma\left(\frac{2n_1+2n_2+i+5}{2}\right) \\ &\times\prod_{i=1}^{L}\left[\Gamma\left(\frac{2n_1+i+1}{2}\right)\Gamma\left(\frac{2n_2+i+1}{2}\right)\Gamma\left(\frac{i}{2}\right)\right]^{-1}. \end{aligned}$$

Explicit expressions of the lower bound for $L = 4$ and 5 are available from Nanda (1947), and for $L = 6$ to 8 from Pillai (1954).

Another special case is when $K = L - 2$. Based on Anderson (1984, p. 305) and Hotelling (1951), the upper bound is:

$$\frac{1-U^{1/2}}{U^{1/2}} \cdot \frac{T-N-1}{N-K} \sim F_{2(N-K),2(T-N-1)}. \tag{26}$$

For additional cases that $K = L - 3$ and $L - 4$, there are also analytical expressions for the upper bound. [4] However, there are no simple formulas for the lower bound.

The testing procedures discussed so far are small sample procedures which are

[4] Following Consul (1966), analytical expressions can be obtained for U that are infinite series of the (Gaussian) hypergeometric functions (Anderson, 1984, p. 113).

valid exactly for any sample size $T \geqslant N + L$. In contrast, the usual approach is to use the asymptotic test that relies on the large sample χ^2 distribution. An alternative and better asymptotic test (Bartlett (1947)) is:

$$\left[T - \tfrac{1}{2}(N + L + 1)\right] \log W_K \overset{\text{asy}}{\sim} \chi^2_{(L-K)(N-K)}. \tag{27}$$

Table 1
A comparison of asymptotic tests with exact test.
The table provides a comparison of the P-values obtained by using the small sample procedure and the asymptotic approximations. It is a case of testing the rank 1 hypothesis in a model where we allow both N and T to vary over a number of values and compute the P-values accordingly

N	T	P-value using exact distribution	P-value using χ^2 approximation	P-value using Bartlett's approximation
10	60	0.050	0.020	0.055
10	120	0.050	0.034	0.053
10	240	0.050	0.042	0.052
20	60	0.050	0.002	0.046
20	120	0.050	0.016	0.054
20	240	0.050	0.030	0.052
30	60	0.050	0.000	0.031
30	120	0.050	0.005	0.050
30	240	0.050	0.021	0.055
40	60	0.050	0.000	0.009
40	120	0.050	0.001	0.043
40	240	0.050	0.011	0.051
50	60	0.050	0.000	0.000
50	120	0.050	0.000	0.034
50	240	0.050	0.005	0.050
100	120	0.050	0.000	0.000
100	240	0.050	0.000	0.030
200	240	0.050	0.000	0.000
10	60	0.100	0.048	0.108
10	120	0.100	0.070	0.102
10	240	0.100	0.087	0.104
20	60	0.100	0.008	0.094
20	120	0.100	0.037	0.100
20	240	0.100	0.066	0.104
30	60	0.100	0.000	0.066
30	120	0.100	0.014	0.098
30	240	0.100	0.044	0.102
40	60	0.100	0.000	0.024
40	120	0.100	0.003	0.088
40	240	0.100	0.027	0.100
50	60	0.100	0.000	0.000
50	120	0.100	0.000	0.072
50	240	0.100	0.013	0.095
100	120	0.100	0.000	0.000
100	240	0.100	0.000	0.064
200	240	0.100	0.000	0.000

Unlike the previous χ^2 approximation, this one has an approximation error that is of order $O(1/T^2)$.

Table 1 provides a comparison of the P-values obtained by using the small sample procedure and by using the asymptotic approximations, respectively. It is a case of testing the rank one hypothesis in a model where $L = 3$, $K = 1$, $\delta_1 = 25$ and $\delta_2 = \delta_3 = 0$. We allow both N and T to vary over a number of values and compute the P-values accordingly. As expected from its high order convergence, Bartlett's asymptotic P-values are consistently better than those from the simple χ^2 approximation. When N is either 10 or 20, there are virtually no differences between the exact P-values and Bartlett's P-values. Although they become better with larger sample size (T), the simple χ^2 P-values are far from accurate. Moreover, neither of the asymptotic approximations is appropriate as the number of assets increases. When $N = 200$, for example, the approximate P-values are less than 0.001 whereas the exact P-value is 10%. Generally speaking, the asymptotic approximations tend to under-estimate the exact P-values, and hence tend to over-reject the null hypothesis if they are used in practice. Notice that the choice of $\delta_1 = 25$ reflects the typical value from our empirical applications (Section 4). If we increase or decrease δ_1 by 10, the exact P-values will change only slightly, similar to the example in the previous subsection. This implies that the asymptotic approximations remain unreliable even with other plausible choices of δ_1. Furthermore, by the monotonicity of the P-value, the asymptotic approximations worsens if a value of δ_1 larger than 25 is used. Therefore, despite the simplicity of the asymptotic approximations, the small sample procedures remain important and may have to be used to make correct statistical inference in applications where the sample size is small compared to the number of assets.

3.3. Equivalence to the likelihood ratio test

The proposed Wald test is in fact equivalent to the usual likelihood ratio test. Such an equivalence was first established by Berndt and Savin (1977) for linear hypotheses and later by Shanken (1985) for testing multi-beta pricing restrictions. It is interesting that there remains a simple linkage between the likelihood ratio test and the Wald test even in the case of the rank hypothesis.

To obtain the ML estimator under the rank constraints, [5] we maximize the

[5] After completion of an earlier version of this paper, Peter Robinson alerted the author to the fact that Anderson (1951), Izenman (1975), Robinson (1973), Velu and Reinsel (1987), and Reinsel (1993), among others, have obtained explicitly the constrained ML estimator. Although partly similar to Tso (1981), our approach computes the estimator efficiently and allows one to obtain the finite sample distribution easily.

likelihood function conditional on Θ and obtain the "concentrated" likelihood function:

$$\log L(\Theta) = -\frac{T}{2}\log|\Omega| + C, \tag{28}$$

where $\Omega = (\mathbf{Y} - \mathbf{X}\Theta)'(\mathbf{Y} - \mathbf{X}\Theta)$ and $C = -NT[\log(2\pi) - \log T + 1]/2$. Under the rank constraints, $\Theta = \mathbf{AB}$ for some $\mathbf{A}$, $L \times K$, and $\mathbf{B}$, $K \times N$. Conditional on $\mathbf{A}$, the estimator of $\mathbf{B}$ is $\tilde{\mathbf{B}} = (\Gamma'\Gamma)^{-1}\Gamma'\mathbf{Y}$ with $\Gamma = \mathbf{XA}$. Hence we need only to find $\mathbf{A}$ that minimizes the 'concentrated' determinant:

$$|\Omega| = |(\mathbf{Y} - \Gamma\tilde{\mathbf{B}})'(\mathbf{Y} - \Gamma\tilde{\mathbf{B}})| = |(\mathbf{Y} - \mathbf{XA}\tilde{B})'(\mathbf{Y} - \mathbf{XA}\tilde{\mathbf{B}})|.$$

If $\mathbf{A}$ and $\mathbf{B}$ are a solution, so are $\mathbf{AU}$ and $\mathbf{U}^{-1}\mathbf{B}$ where $\mathbf{U}$ is an arbitrary nonsingular $K \times K$ matrix. Thus we can assume, without the loss of generality, that $\Gamma'\Gamma = \mathbf{A}'\mathbf{X}'\mathbf{XA} = \mathbf{I}_K$. Then $|\Omega|$ can be written:

$$|\Omega| = |\mathbf{Y}'\mathbf{Y}| \cdot |(\mathbf{QA})'(\mathbf{I}_L - \mathbf{Z})(\mathbf{QA})|,$$

where both $\mathbf{Q}$ and $\mathbf{Z}$ are $L \times L$ matrices:

$$\mathbf{Z} \equiv \mathbf{Q}^{-1\prime}\mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{XQ}^{-1},\ \mathbf{Q}'\mathbf{Q} \equiv \mathbf{X}'\mathbf{X}. \tag{29}$$

Hence, we need only to minimize $|(\mathbf{QA})(\mathbf{I}_L - \mathbf{Z})(\mathbf{QA})|$ which can be done by using the *Poincare Separation Theorem*:

Lemma 1. Let $\mathbf{V}$ be an $L \times K$ matrix such that $\mathbf{V}'\mathbf{V} = \mathbf{I}_K$, and $\mathbf{U}$ an $L \times L$ positive definite matrix. Then the i-th largest eigenvalue of $\mathbf{V}'\mathbf{UV}$ is bounded both below and above by the eigenvalues of $\mathbf{U}$:

$$\lambda_{L-K+i}(\mathbf{U}) \leqslant \lambda_i(\mathbf{V}'\mathbf{UV}) \leqslant \lambda_i(\mathbf{U}),\ i = 1, \ldots, K. \tag{30}$$

Furthermore, the right-hand side becomes an equality if and only if V is formed by the eigenvectors corresponding to the K largest eigenvalues of $\mathbf{U}$, and the left-hand side becomes an equality if and only if $\mathbf{V}$ is formed by the eigenvectors corresponding to the K smallest eigenvalues.

Proof. Rao (1973, P. 64). **Q.E.D.**

As an immediate consequence of this Lemma, we have $|\mathbf{V}'\mathbf{UV}| \geqslant \lambda_{L-K+1}(\mathbf{U}) \cdots \lambda_L(\mathbf{U})$, implying that the minimum of $|\Omega|$ is a product of $|\mathbf{Y}'\mathbf{Y}|$ and the K smallest eigenvalues of $\mathbf{I}_L - \mathbf{Z}$. Let $\lambda_1 \geqslant \ldots \geqslant \lambda_L \geqslant 0$ be the eigenvalues of $\mathbf{Z}$, and $\mathbf{P}$ a matrix of the eigenvectors, then $|\Omega|$ is minimized if $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{P}$ and the minimum is given by:

$$|\Omega_r| = |\mathbf{Y}'\mathbf{Y}|(1 - \lambda_1) \cdots (1 - \lambda_K). \tag{31}$$

A similar formula for the unconstrained $|\Omega|$ is:

$$|\Omega_u| = |\mathbf{Y}'\mathbf{Y}|(1 - \lambda_1)(1 - \lambda_2) \cdots (1 - \lambda_L). \tag{32}$$

A combination of (31) and (32) yields a formula for the likelihood ratio:

$$LR = \left[(1 - \lambda_{K+1}) \cdots (1 - \lambda_L)\right]^{T/2}. \tag{33}$$

Equations (31)–(33) are well-known results (see, e.g., Tso, 1981). They are derived here by a slightly shorter method, and are provided for the completeness of the paper.

To summarize, the constrained maximum likelihood estimator of Θ and Σ is:

$$\tilde{\Theta} = \mathbf{Q}^{-1}\mathbf{P}\mathbf{P}'\mathbf{Q}^{-1\prime}\mathbf{X}'\mathbf{Y}, \text{ and } \tilde{\Sigma} = \frac{1}{T}(\mathbf{Y} - \mathbf{X}\tilde{\Theta})'(\mathbf{Y} - \mathbf{X}\tilde{\Theta}), \tag{34}$$

where $\mathbf{P}$, $L \times K$, is the matrix formed by the 'standardized' eigenvectors ($\mathbf{P}'\mathbf{P} = \mathbf{I}_K$) corresponding to the first K eigenvalues of $\mathbf{Z}$ from (29). There are only two matrix inversions to be performed, the $L \times L$ and $N \times N$ matrices $\mathbf{Q}^{-1}$ and $(\mathbf{Y}'\mathbf{Y})^{-1}$ which are easy to compute in most applications. Since $\mathbf{P}\mathbf{P}'$ is invariant to permutations of the columns of $\mathbf{P}$, the constrained estimator is unique.

By the likelihood ratio principle, LR should be close to 1 if the null hypothesis (2) is true. In this case, we expect to observe small $\lambda_{K+1}, \ldots, \lambda_L$. Any function of the eigenvalues, such as the maximum of $\lambda_{K+1}, \ldots, \lambda_L$, can be used as a test statistic. When the maximum is small, LR should be close to 1 and we cannot reject the null. However, we use in what follows the W_K statistic:

$$W_K \equiv \prod_{j=K+1}^{L} \frac{1}{1-\lambda_j}, \tag{35}$$

for two simple reasons. First, it is the likelihood ratio test [$\log W_K = -(2/T)\log(LR)$]. Second, its distribution has simple analytical bounds for some useful special cases, providing additional insights on the test and on the use of the Monte Carlo integration approach. [6]

Finally, we want to show that the Wald test is equivalent to the likelihood ratio test. Because $\hat{\mathbf{W}}$ from (18) and $\mathbf{Z}$ from (29) are not identical matrices, we cannot in general expect that their eigenvalues are related to one another. Interestingly though, it can be shown that (see Appendix A):

$$\lambda_j = \xi_j/(1+\xi_j), \; j = 1, \ldots, L. \tag{36}$$

Hence, the W_K defined by (19) must be identical to that defined by (35), i.e., the proposed Wald test is equivalent to the likelihood ratio test.

4. Empirical results

In this section, we apply our methodology to examine the number of latent factors in the US equity market. Following Ferson and Harvey (1991) and others, we get twelve industry returns (petroleum, finance/real estate, consumer durables,

[6] An additional test, $W_2 = \sum_{j=K+1}^{L} \lambda_j/(1-\lambda_j)$, also has simple analytical bounds. This test is an analogue of the Lawley–Hotelling's trace test for linear hypotheses.

Table 2
Means, standard deviations autocorrelations of monthly portfolio returns from October 1941 to September 1986 (540 observations)

Industry portfolio returns [a]								
Variable	Mean	Std. dev.	Autocorrelation					
			ρ_1	ρ_2	ρ_3	ρ_4	ρ_{12}	ρ_{24}
Industry 1	1.199	5.005	0.011	−0.030	0.049	0.068	0.031	0.000
Industry 2	1.091	4.426	0.061	−0.013	−0.007	0.038	0.019	0.014
Industry 3	1.115	5.023	0.078	0.025	0.022	0.054	0.010	−0.018
Industry 4	1.001	4.488	0.020	0.000	0.030	0.072	0.016	0.013
Industry 5	1.056	3.855	0.117	0.018	0.040	0.054	0.036	−0.029
Industry 6	0.960	5.195	0.113	−0.026	−0.006	0.026	0.022	0.026
Industry 7	1.107	4.773	0.103	−0.007	0.038	0.047	0.041	0.045
Industry 8	1.043	5.692	0.071	0.027	−0.021	0.099	0.031	0.028
Industry 9	0.921	3.336	0.105	−0.014	0.039	0.092	0.039	0.053
Industry 10	1.047	4.860	0.132	0.052	0.010	0.053	0.046	−0.070
Industry 11	1.247	5.688	0.131	0.042	0.023	0.043	0.034	−0.031
Industry 12	1.256	5.757	0.208	0.090	0.034	0.000	0.047	−0.104

[a] The industry groups are: 1 = petroleum, 2 = finance/real estate, 3 = consumer durables, 4 = basic industries, 5 = food/tobacco, 6 = construction, 7 = capital goods, 8 = transportation, 9 = utilities, 10 = textiles/trade, 11 = services 12 = leisure.

basic industries, food/tobacco, construction, capital goods, transportation, utilities, textiles/trade, services and leisure). By arbitrarily choosing petroleum as the first asset, eleven excess returns are obtained as the returns in excess of the return on petroleum. The returns are monthly data from October 1941 to September 1986 (45 years). Although the sample size T can be any number provided $T \geqslant N + L$, the parameter estimation may not be very accurate if T is too small. On the other hand, concerns about parameter stability suggest that T should not be too large. Following the usual practices, we choose the five-year subperiods. For the sake of comparison, we also provide results for the entire sample period.

Means, standard deviations and autocorrelations of the data are presented in Table 2. The means range from 0.92% per month (11.04% per year) for the utilities industry (industry 9) to 1.26% per month for the leisure industry (industry 12). The lowest standard deviation, a measure of the total industry risk, is found in the utilities industry and the highest is found in the leisure industry. Although for both industries 9 and 12, the high or low average returns are associated with their industry risks, the petroleum industry (industry 1) has lower risk and higher return than the transportation industry (industry 8). However, this is not in contradiction with financial theories that high returns should be associated with high systematic risks which are not necessarily the total risk as measured by the standard deviation. There is only very small first order autocorrelations and others die out fast as the lags increase.

There are two sets of instrumental variables. The first is a *small set* which contains returns on the value-weighted and equal-weighted indices plus a constant. The second is a *large set* comprising: the returns on the equal-weighted index in excess of the 30-day Treasury bill rate, a dummy variable for January, the return for holding a 90-day bill for one month less the return on the 30-day bill (term premium), the yield on Moody's BAA-rated bonds less the yield on Moody's AAA-rated bonds (junk bond premium), the dividend yield on the Standard and Poor's 500 stock index less the return on a 30-day bill (dividend yield spread), and a constant term. To be consistent with the model assumption that investors know the instrumental variables at the time when forming their expectations about the stock returns, all the instrumental variables of returns are lagged one month.

Although the number of instrumental variables L can be arbitrary provided $K < L < N$, the small set is chosen because it contains market information as reflected by the indices. The large set is chosen to provide a wide selection of variables that are potentially those investors may use to forecast stock returns. The predictability of the term premium on stock returns was established by Fama (1984) and Campbell (1987), that of the dividend yield spread by Fama and French (1988, 1989) and Campbell and Shiller (1988), and that of the junk bond premium by Keim and Stambaugh (1986).

Table 3 summarizes the results of testing whether there is only one 'priced' latent factor. Panel A provides the results over each five-year subperiod of the sample. The second column is the Wald test statistic (W_K) and the third is the associated exact P-value (P_W), both of which are the results in the case where the small instrument set is used. The P-values vary from a low value of 0.169 to a high value of 0.965. We cannot reject the null hypothesis at the usual 5% level, or even at the 10% level. To study the sensitivity to the choice of the instrumental variables, we test the one factor hypothesis again by employing the large instrument set which contains more information variables, such as the junk bond premium (the credit change of companies), than the small instrument set. The results are reported in the fourth and fifth columns of the table. At the 5% level, there is only one P-value, 0.030 for the period 1961/10–1966/9, that is slightly below 5%. For all other periods, the P-values are greater than 5%. Overall, there is almost no evidence at all against the null hypothesis that there is one latent factor versus the alternative hypothesis that there are at least two factors.

An alternative procedure is Hansen's (1982) GMM test. As pointed out earlier, this test cannot be used in our studies over the five-year subperiods due to the singularity of the optimal weighting matrix. Over the entire period, the test can be used, and is obtained analytically based on Zhou (1994). Both the test statistic and the associated asymptotic P-value are reported in Panel B of Table 3. The P-values from the GMM test are 85% and 6.6% for the two instrument sets, respectively. We cannot again reject a one-factor model. We also report in Panel B the evidence of the Wald test over the entire sample period. When the small instrument set is used, the P-value is 52.9%, favoring a one-factor model.

Table 3
Tests of one latent factor for stock returns
We test whether the expected industry excess returns are driven by only one latent factor:
$E(R_{it} | Z_{t-1}) = b_{i1}\lambda_1(\mathbf{Z}_{t-1})$, $i = 1, \ldots, N$,
where $\lambda_1(\mathbf{Z}_{t-1})$ is the factor risk premium, $\mathbf{Z}_{t-1}$ is the information set b_{i1} is the excess conditional beta. Panel A provides the results over each five-year subperiod of the sample. The second column is the Wald-type test statistic (W_K) the third is the associated exact P-value (P_W). The fourth fifth are similar results except that the large instrument set is used instead of the the small instrument set. Panel B provides the results of using both the Wald-type test the GMM test for the whole period

Period	Panel A: Five-year subperiods ($T = 60$)			
	Small instrument set		Large instrument set	
	W_K	P_W	W_K	P_W
1941/10–1946/9	1.6415	0.169	2.8454	0.356
1946/10–1951/9	1.3613	0.703	3.3242	0.137
1951/10–1956/9	1.3816	0.655	2.8912	0.325
1956/10–1961/9	1.4270	0.544	3.5660	0.081
1961/10–1966/9	1.5401	0.308	3.9997	0.030
1966/10–1971/9	1.5023	0.375	2.9972	0.266
1971/10–1976/9	1.2144	0.965	2.5562	0.565
1976/10–1981/9	1.3578	0.713	3.6727	0.064
1981/10–1986/9	1.3154	0.808	3.2691	0.155
	Panel B: Whole period ($T = 540$)			
Wald-type test	1.0359	0.529	1.1768	0.002
GMM test	13.567	0.850	65.857	0.066

However, the P-value is only 0.2% when the large instrument set is used. Concerns of parameter stability over the entire sample period suggest that we should give more weight to the subperiod results. So, we interpret the overall evidence as in favor of a one-factor model. [7]

The fact that we cannot reject a one-factor model may not be disappointing. In fact, many asset pricing theories are one-factor models and the difference among the models is in the specification of the factor. For examples, the classical capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) and the zero-beta CAPM of Black (1972) are one-factor models, which identify the factor as the market portfolio. The arbitrage pricing theory (APT) of Ross (1976) has a L-factor returns generating process but under certain equilibrium conditions, the return on assets are rewarded by only one factor which is a portfolio of the L-factors in the returns generating process. In a multi-good, continuous-time model with uncertain consumption-goods prices and uncertain investment opportunities, Breeden (1979) shows that the multi-beta asset pricing model of Merton

[7] To obtain the analytical GMM test in Table 3, we use theorem 2 of Zhou (1994) first, and then Theorem 1 there with $d = NL\text{-}\delta$, where $\delta = K(N - K + L)$.

(1973) collapses into a single-beta model with aggregate consumption as the only 'priced' latent factor. Therefore, one may interpret the evidence here as general support of one-factor asset pricing theories. But it does not, of course, discriminate among them.

5. Conclusions

We propose small sample tests for rank restrictions on the regression coefficient matrix in the classical multivariate regression model. As an application, we investigate the number of 'priced' latent factors in the US equity market by using monthly portfolio returns grouped by industry. Based on two sets of instrumental variables, we cannot reject the hypothesis that a one-factor model explains the industry returns reasonably well.

There are many issues that appear of great interest for further research. It would be interesting to apply the small sample results to study stock returns, term structure theories, forward currency premiums, international equity returns and capital integration, and the reduction of factors in a variety of contexts, and to compare the results with those obtained by the asymptotic GMM approach. In addition, because both the estimation and test procedures proposed in this paper rely on the normality assumption, further studies are needed to assess the robustness to non-normality. In particular, following Rothenberg (1984), it may be useful to develop an Edgeworth expansion for the tests in the time series case, complementing earlier studies by Robinson (1973) and Velu and Reinsel (1987). Finally, a Bayesian posterior analysis of the deltas and an odds-ratio approach for testing the rank constraints seem to be an important topic to pursue. This is likely to be more appealing in the case where the disturbances are Gaussian VARMA. For the VARMA models, it appears extremely difficult to obtain any small sample results, but an exact Bayesian inference may be feasible by using Monte Carlo integration with importance sampling (Geweke (1989)) or by using the Gibbs sampling approach (Geweke (1994)).

Appendix A: *Proofs*

1. Proof of Theorem 1:

Although the estimation technique of Zhou (1991) is not applicable in the $K > 1$ case, the proof can be easily extended to prove Theorem 1 based on the relationship between the eigenvalues of $\hat{\mathbf{W}}$ from (18) and $\mathbf{Z}$ from (29). However, we provide in what follows an alternative and rather simple proof.

It is well known from multivariate analysis (e.g., Muirhead, 1982, p. 431) that $\hat{\Theta}$ and $\hat{\Sigma}$ are statistically independent and have multivariate normal and Wishart

distributions respectively (using the matrix normal notation of Muirhead, 1982, p. 79):

$$\hat{\Theta} \sim N\big(\Theta, (\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma\big),\ T\hat{\Sigma} \sim W_N(T-L, \Sigma).$$

It follows that:

$$\mathbf{Y}_1^* \equiv (\mathbf{X}'\mathbf{X})^{1/2}\hat{\Theta} \sim N\big((\mathbf{X}'\mathbf{X})^{1/2}\Theta, \mathbf{I}_L \otimes \Sigma\big).$$

By (18), $\xi_1, \ldots, \xi_L$ are eigenvalues of the matrix $\mathbf{Y}_1^*(\mathrm{T}\hat{\Sigma})^{-1}\mathbf{Y}_1^{*\prime}$.

Now notice that $(T-L) \geqslant N \geqslant L$, we have from Theorem 10.4.5 of Muirhead (1982, p. 454) that the joint density function of $\xi_1, \ldots, \xi_L$ is given by:

$$f(\xi_1, \ldots, \xi_L) = etr\big(-\tfrac{1}{2}\Delta\big)\,{}_1F_1\left(\frac{T}{2}; \frac{N}{2}; \frac{\Delta}{2}, \mathbf{F}(\mathbf{I}+\mathbf{F})^{-1}\right) \times \frac{\pi^{L2/2}\Gamma_L(T/2)}{\Gamma_L(L/2)\,\Gamma_L(N/2)\,\Gamma_L((T-N)/2)} \prod_{i=1}^{L} \frac{\xi_i^{n_1}}{(1+\xi_i)^{n_2}} \prod_{i<j}^{L} (\xi_i - \xi_j),$$

where $\Delta = \mathbf{M}_1\Sigma^{-1}\mathbf{M}_1'$, $\mathbf{M}_1 = (\mathbf{X}'\mathbf{X})^{1/2}\Theta$, $\mathbf{F} = \mathrm{diag}(\xi_1, \ldots, \xi_L)$, $n_1 = (N-L-1)/2$, $n_2 = (T-N-L-1)/2$, *etr* is the exponential function of the trace operator, and ${}_1F_1(a; b; \mathbf{M}_1; \mathbf{M}_2)$ is the hypergeometric function of matrix arguments (see James, 1964 or Chapter 7 of Muirhead, 1982). Because $f(\xi_1, \ldots, \xi_L)$ depends on Δ only through its eigenvalues (see (20)), we can write $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_L)$.

Finally, since $\mathbf{Y}_1^*$ is $L \times N$, $\mathbf{A}^* \equiv \mathbf{Y}_1^{*\prime}\mathbf{Y}_1^*$ will be singular when $N > L$. Let $\mathbf{B}^* \equiv T\hat{\Sigma}$. It is seen that $\xi_1, \ldots, \xi_L$ are non-zero eigenvalues of $\mathbf{A}^*\mathbf{B}^{*-1}$, a "ratio" of a possibly singular Wishart matrix to a non-singular Wishart matrix. By the important results for singular Wishart matrix developed first by James (1964), the non-zero eigenvalues of $\mathbf{A}^*\mathbf{B}^{*-1}$ are equivalent to those of two non-singular Wishart matrices with the adjusted degrees of freedom as stated in Theorem 1 (see the comments on p. 455 of Muirhead (1982) and Theorem 10.4.2 there). **Q.E.D.**

2. Proof of equation (36):

By definition, $\xi_1 \geqslant \xi_2 \geqslant \cdots \geqslant \xi_L \geqslant 0$ are the roots of the determinantal equation:

$$\left|(\mathbf{X}'\mathbf{X})^{1/2}\hat{\Theta}\left[(\mathbf{Y}-\mathbf{X}\hat{\Theta})'(\mathbf{Y}-\mathbf{X}\hat{\Theta})\right]^{-1}\hat{\Theta}'(\mathbf{X}'\mathbf{X})^{1/2} - \xi\mathbf{I}\right| = 0$$

Using the equality that $|\mathbf{UV} - \xi\mathbf{I}_L| = |\mathbf{VU} - \xi\mathbf{I}_N|$ (see, e.g., Zellner, 1971, p. 231), we have, after multiplying $|T\hat{\Sigma}|$ on the lefthand side,

$$\left|\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \xi\,\mathbf{Y}'\left[\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{Y}\right| = 0.$$

Combining the terms, dividing the result by $(1 + \xi)$, we get:

$$|\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - [\xi/(1+\xi)]\mathbf{Y}'\mathbf{Y}| = 0.$$

Now, multiplying by $(\mathbf{Y}'\mathbf{Y})^{-1}$ on the both sides and using again the equality that $|\mathbf{UV} - \xi\mathbf{I}_L| = |\mathbf{VU} - \xi\mathbf{I}_N|$, we obtain exactly the determinantal equation for the eigenvalues of $\mathbf{Z}$ with $\xi/(1+\xi)$ plays the role of λ. Therefore we have $\lambda_j = \xi_j/(1+\xi_j)$, $j = 1, \ldots, L$. **Q.E.D.**

Appendix B: *Drawing samples from Wishart distributions*

1. Drawing samples from a central Wishart distribution

Based on Bartlett's decomposition (Muirhead (1982), p. 99), we have a two-step procedure: (i) draw independent $L_{ij} \sim N(0, 1)$, $i > j$ and $L_{ii}^2 \sim \chi^2(N - i + 1)$ to form a lower triangular matrix $\mathbf{L}$; (ii) compute the $m \times m$ matrix: $\mathbf{S} \equiv \mathbf{LL}'$, then $\mathbf{S} \sim W_m(N, \mathbf{I}_m)$. In general, to draw samples from $W_m(N, \Sigma)$, we obtain first the LU decomposition of Σ, $\Sigma = \mathbf{PP}'$. Then $\mathbf{PSP}' \sim W_m(N, \Sigma)$. Notice that if a number of samples are to be generated, $\mathbf{P}$ needs to be computed only once.

2. Drawing samples from a non-central Wishart distribution

By a theorem due to Anderson (Johnson and Kotz (1972), p. 177), we have: (i) draw a Wishart sample $\mathbf{S} \sim W_m(N - m, \mathbf{I}_m)$; (ii) draw m^2 independent normal samples $z_{ij} \sim N(0, 1)$ to form a $m \times m$ matrix $\mathbf{Z}$; (iii) compute the $m \times m$ matrix: $\mathbf{V} = \mathbf{S} + (\mathbf{Z} + \mathbf{M})(\mathbf{Z} + \mathbf{M})'$, then $\mathbf{V} \sim W_m(N, \mathbf{I}_m, \mathbf{M})$.

References

Anderson, T.W., 1951, Estimating linear restrictions on regression coefficients for multivariate normal distribution, Annals of Mathematical Statistics 29, 813–828.

Anderson, T.W., 1984, An introduction of multivariate statistical analysis, second edition (Wiley, New York, NY).

Bartlett, M.S., 1947, Multivariate analysis, Journal of the Royal Statistical Society (Suppl.) 9, 176–190.

Berndt, E. and N. Savin, 1977, Conflict among criteria for testing hypotheses in the multivariate regression model, Econometrica 45, 1263–1278.

Black, F., 1972, Capital market equilibrium with restricted borrowing, Journal of Business 45, 444–454.

Breeden, D.T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, Journal of Financial Economics 7, 265–296.

Campbell, J.Y., 1987, Stock returns and the term structure, Journal of Financial Economics 18, 373–399.

Campbell, J.Y. and R.J. Shiller, 1988, The dividend price ratio and expectations of future dividends and discount factors, Review of Financial Studies 1, 195–228.

Chen, C-F., 1981, The EM approach to the multiple indicators and multiple causes model via the estimation of the latent variable, Journal of the American Statistical Association 76, 704–708.

Consul, P.C., 1966, On the exact distribution of the likelihood ratio criteria for testing linear hypotheses about regression coefficients, Annals of Mathematical Statistics 37, 1319–1330.

Cox, J.C., J.E. Ingersoll and S.A. Ross, 1985, A theory of the term structure of interest rates, Econometrica 53, 385–407.

Davies, P.T. and M.K.S. Tso, 1982, Procedures for reduced rank regression, Applied Statistics 31, 244–255.

Fama, E.F., 1984, The information in the term structure, Journal of Financial Economics, 13, 509–528.

Fama, E.F. and K.R. French, 1988, Dividend yields and expected stock returns, Journal of Financial Economics, 22, 3–25.

Fama, E.F. and K.R. French, 1989, Business conditions and expected stock returns, Journal of Financial Economics, 25, 23–50.

Ferson, W.E. and C.R. Harvey, 1991, The variation of economic risk premiums, Journal of Political Economy 99, 385–415.

Geweke, J.F., 1989, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57, 1317–1339.

Geweke, J.F., 1994, Bayesian reduced rank regression in econometrics, Journal of Econometrics, forthcoming.

Gibbons, M.R., 1982, Multivariate tests of financial models: A new approach, Journal of Financial Economics 10, 3–27.

Gibbons, M.R. and W. Ferson, 1985, Testing asset pricing models with changing expectations and an unobservable market portfolio, Journal of Financial Economics 14, 217–236.

Hansen, L.P., 1982, Large sample properties of the generalized method of moments estimators, Econometrica 50, 1029–1054.

Hansen, L.P., and R.J. Hodrick, 1983, Risk averse speculation in the forward foreign exchange market: an econometric analysis of linear models, in: J.A. Frenkel, ed., Exchange Rates and International Macroeconomics (University of Chicago Press, Chicago, IL), 113–152.

Hotelling, H., 1951, A generalized T test and measure of multivariate dispersion, in: J. Neyman, ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California, Los Angels and Berkeley), 111–184.

Izenman, A.J., 1975, Reduced-rank regression for the multivariate linear model, Journal of multivariate analysis 5, 248–264.

James, A.T., 1964, Distributions of matrix variates and latent roots derived from normal samples, Annals of Mathematical Statistics 35, 475–501.

Johnson, N.J., and S. Kotz, 1972, Distributions in statistics: continuous multivariate distributions (Wiley, New York, NY).

Jöreskog, K.G. and A.S. Goldberger (1975), Estimation of a model with multiple indicators and multiple causes of a single latent variable, Journal of the American Statistical Association 70, 631–639.

Keim, D.B. and R.F. Stambaugh, 1986, Predicting returns in the bond and stock market, Journal of Financial Economics, 17, 357–390.

Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics 47, 13–37.

Merton, R.C., 1973, An intertemporal capital asset pricing model, Econometrica 41, 867–887.

Muirhead, R.J., 1982, Aspects of multivariate statistical theory (Wiley, New York, NY).

Nanda, D.N., 1947, Distribution of a root of a determinantal equation, Annals of Mathematical Statistics 18, 47–57.

Perlman, M.D. and I. Oklin, 1980, Unbiasedness of invariant tests for MANOVA and other multivariate problems, Annals of Statistics 8, 1326–1341.

Pillai, K.C.S., 1954, On some distribution problems in multivariate analysis (Memeo Series no. 88, Institute of Statistics, University of North Carolina).

Rao, C.R., 1973, Linear statistical inference and its applications, second Edition (Wiley, New York, NY).

Reinsel, G.C. 1993, Elements of multivariate time series analysis (Springer-Verlag, New York, NY).

Robinson, P.M., 1973, Generalized canonical analysis for time series, Journal of Multivariate Analysis 3, 141–160.

Ross, S.A., 1976, The arbitrage theory of capital asset pricing, Journal of Economic Theory 13, 341–360.

Rothenberg, T.J., 1984, Approximating the distributions of econometric estimators and test statistics, in: Z. Griliches and M.D. Intriligator, ed., Handbook of Econometrics, II (North-Holland, Amsterdam: North Holland), 882–935.

Sargent, T.J., and C.A. Sims, 1977, Business cycle modeling without pretending to have too much of a prior economic theory, in: C.A. Sims, ed., New methods in business cycle research, Proceedings from a conference (Federal Reserve Bank of Minneapolis, MN), 45–110.

Schott, J.R., 1984, Optimal bounds for the distributions of some test criteria for tests of dimensionality, Biometrika 71, 561–567.

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

Shanken, J., 1986, Testing portfolio efficiency when the zero-beta rate is unknown: A note, Journal of Finance 41, 269–276.

Sharpe, W.F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 19, 425–442.

Stambaugh, R.F., 1988, The information in forward rates: Implications for models of the term structure, Journal of Financial Economics 21, 41–70.

Tso, M.K.S, 1981, Reduced rank regression and canonical analysis, Journal of the Royal Statistical Society B 43, 183–189.

Velu, R.P., 1992, Reduced rank models with two sets of regressors, Applied Statistics 40, 159–170.

Velu, R.P. and G.C. Reinsel, 1987, Reduced rank regression with autoregressive errors, Journal of Econometrics 35, 312–335.

Zellner, A, 1970, Estimation of regression relationships containing unobservable variables, International Economic Review 11, 441–454.

Zellner, A, 1971, An introduction to Bayesian inference in econometrics (Wiley, New York, NY).

Zhou, G., 1991, Small sample tests of portfolio efficiency, Journal of Financial Economics 30, 165–191.

Zhou, G., 1994, Analytical GMM tests: Asset pricing with time-varying risk premiums, Review of Financial Studies 7, 687–709.