GMM Estimation of the Number of Latent Factors

Perez, Marcos and Ahn, Seung Chan
Arizona State UNiversity

09. September 2007

GMM Estimation of the Number of Latent Factors

Seung Chan Ahn

Marcos Fabricio Perez

W.P. Carey School of Business
Arizona State University
Tempe, AZ 85287-3906

This Version: September 9, 2007

GMM Estimation of the Number of Latent Factors

Abstract

We propose a generalized method of moment (GMM) estimator of the number of latent factors in linear factor models. The method is appropriate for panels a large (small) number of cross-section observations and a small (large) number of time-series observations. It is robust to heteroskedasticity and time series autocorrelation of the idiosyncratic components. All necessary procedures are similar to three stage least squares, so they are computationally easy to use. In addition, the method can be used to determine what observable variables are correlated with the latent factors without estimating them. Our Monte Carlo experiments show that the proposed estimator has good finite-sample properties. As an application of the method, we estimate the number of factors in the US stock market. Our results indicate that the US stock returns are explained by three factors. One of the three latent factors is not captured by the factors proposed by Chen Roll and Ross 1986 and Fama and French 1996.

Keywords: Factor models, GMM, number of factors, asset pricing.

1. INTRODUCTION

Many economic and financial theories are based on linear factor models. A well known example is the Arbitrage Price Theory (APT, Ross, 1976), where asset returns are generated by a factor structure. In the finance literature, the APT model has been extensively used to analyze the prices of the systematic risks in the stock, money, or fixed income securities markets. There are many other examples. Analyzing the data from G7 countries, Gregory and Head (1999) found that cross-country variations in productivity and investment have common components. Gorman (1981) and Lewbel (1991) found that if consumers are utility maximizers, their budget shares for individual goods or services purchased should be driven by at most three factors. Stock and Watson (2005) proved that many macroeconomic variables in US are driven by a smaller number of common factors. Ahn, Lee and Schmidt (2007a) showed that the time pattern of the fluctuations in individual firms' technical productivities can be estimated based on a factor model. An excellent summary of the use of factor models can be found in Campbell, Lo and Mackinlay (1997) and also in Bai (2003).

For any empirical study that involves factor models, estimation of the true number of factors is crucial in order to identify and estimate the factors. It is also important to determine what observable macroeconomic and/or financial variables are related to the unobservable factors, in order to give an economic interpretation to the model. We propose a methodology to address these questions using an estimation procedure based on GMM.

Earlier empirical studies of factor models were based on the maximum likelihood (ML) method of Jöreskog (1967). Using this method, a researcher estimates factor loadings and variances of idiosyncratic errors of asset returns concurrently, and test for the number of latent factors using a likelihood-ratio test. The ML method requires quite restrictive distributional assumptions: the idiosyncratic error terms are required to be normal, and independently and identically distributed over time. More general approaches have been developed allowing for less restrictive assumptions. A common method is to construct candidate factors, repeat the estimation and testing of the model for different number of factors (L), and observe if the tests are sensitive to increasing L. Lehman & Modest (1988) and Connor & Korajczyk (1988) used this technique to analyze the US stock returns. Success of this method would depend on the quality of the chosen candidate factors. Another approach is to use estimators of the ranks of matrices[1] (*e.g.*, Gill and Lewbel, 1992; Cragg and Donald, 1996, 1997). A limitation of this approach is that it is computationally burdensome, especially if the number of response variables analyzed is large.[2] More recently, Bai and Ng (2002) have developed a general estimation method for the number of factors. Their least squares estimation method is designed for data with a large number of response variables (N) and a large number of time series observations (T). This method could produce inconsistent estimators if either N or T is small. Simulation results reported in Bai and Ng (2002) indicate that the number of factors is not accurately estimated if N or T is less than 40. Thus, the least squares method would be inappropriate for the studies using small sets of response variables.

In this paper we present an alternative generalized method of moment (GMM) estimator of the number of factors. The advantages of this new method compared with those discussed above are the following. First, the method requires that just one of the data dimensions (N or T) to be large; that is, either the number of cross-section or time series observations has to be large. Several economic and financial applications involve small cross sectional observations. Examples are the analyses of portfolio returns, yields on bond indexes, or country common factors. Second, the method provides a way to check possible correlations between observable variables (i.e., macroeconomics or financial variables) and unobservable factors without estimating factor themselves. Using our method, researchers are able to give an economic interpretation to the latent factors model (see, Ahn, Dieckmann and Perez, 2007). Third, the method is computationally easy to implement. All necessary procedures are based on closed-form solutions, and thus, do not require non-linear optimization. Any software that can estimate multiple equations models can be used. Fourth, the method allows for cross-section and time series heteroskedasticity and time series autocorrelation of the idiosyncratic components and it

---

[1] If the idiosyncratic error components of the response variables analyzed are cross-sectionally independent (exact factor model), the variance matrix of the response variables (*e.g.*, returns) is decomposed into a diagonal matrix and a matrix with a rank equal to L. Thus, the number of the common factors (L) can be found by estimating the rank of the difference between the estimates of the variance and the diagonal matrices.

[2] Rank of a matrix can be estimated by the Lower-Diagonal-Upper triangular decomposition test (LDU) developed by Gill and Lewbel (1992) and Cragg and Donald (1996). This method requires a Gaussian elimination procedure and division of the response variables into two non-overlapping groups. The Gaussian elimination procedure is complicated if too big matrices are analyzed. Alternatively, Cragg and Donald (1997) propose a Minimum Chi-Squared statistic (MINCHI2). This method is general in the sense that it requires only weak distributional assumption about the response variables and allows for heteroskedasticity and autocorrelation. The principal problem of MINCHI2 is that some nonlinear optimization procedures are required and the procedures often fail to locate solutions as shown by Donald, Fortuna and Pipiras (2005).

does not require distributional assumptions about the data generating process. Our method is primarily designed for exact factor models in which idiosyncratic error components of response variables are cross-sectionally uncorrelated. However, even if the errors are cross sectionally correlated, the method can be used to estimate the number of factors if N is large and the response variables can be grouped appropriately (e.g., portfolios).

As an application we use our methodology to analyze the US stock market. Our empirical results imply that the US stock returns are determined by three factors. Also we find that the variables proposed by Chen Roll and Ross 1986 are able to capture just one of the three latent factors. Fama and French 1996 proposed factors are able to capture an extra latent factor. One of the three unobservable factors is not captured by the factors proposed by Chen Roll and Ross 1986 and Fama and French 1996.

The rest of the paper is organized as follows. Section 2 introduces the factor model we investigate, and lists the basic assumptions we made for the estimation. Section 3 explains our GMM method to estimate the number of factors. In section 4, we consider how the method could be used for the analysis of the models when the idiosyncratic components are cross-sectionally correlated. We also consider the cases in which some observable variables that are potentially correlated with latent factors. Section 5 exhibits our Monte Carlo simulation results and finite-sample properties of our method. Section 6 discusses the results we obtain by applying the method to the U.S. stock market. Concluding remarks are provided in Section 7.

2. Model and Assumptions

We consider a linear model with a finite number of unobservable latent common factors:

$$r_{it} = \alpha_i + \beta_i' f_t + \varepsilon_{it} , \tag{1}$$

where r_{it} is the value of the response variable i (= 1, 2, …, N) at the time t (= 1, 2, …, T), α_i is an intercept, f_t is an $L \times 1$ vector of unobservable common factors, β_i is an $1 \times L$ vector of the factor loadings for the response variable i, and the ε_{it} are the idiosyncratic components of response variables which are cross-sectionally uncorrelated. Thus, the response variables r_{it} are cross-sectionally correlated only through the common factors f_t. Usual factor analysis typically applies to demeaned data with $E(r_{it}) = 0$ for all *i* and *t*. But we do not impose such restrictions. To begin, we consider the cases in which N is relatively small and T is large. Thus, the asymptotic theory we use below applies as $N \to \infty$ for fixed T. We will consider later the cases in which T is large and N is small.

For convenience, we adopt the following notation. We use $r_{\bullet t}$ to denote the vector that includes all the cross-sectional observations of the response variable r_{it} at time *t*. Similarly, $r_{i\bullet}$ denotes the vector including all of the time series observations of r_{it} for the response variable *i*. The vectors $\varepsilon_{i\bullet}$ and $\varepsilon_{\bullet t}$ are similarly defined. Using this notation, we can stack the equations in (1) for given t by

$$r_{\bullet t} = \alpha + \mathrm{B} f_t + \varepsilon_{\bullet t}, \tag{2}$$

where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)'$ and $\mathrm{B} = (\beta_1, \beta_2, ..., \beta_N)'$. Including the non-zero vector of response-variable-specific intercepts into the model, we can assume that $E(f_t) = 0$ without loss of generality.

Our method to estimate the number of factors (*L*) is an application of GMM. Thus, we require a set of sufficient conditions under which usual GMM theories apply and the number of factors can be identified. For asymptotics, we use "$\to_p$" and "$\to_d$" to denote "converges in probability" and "converges in distribution," respectively. The basic assumptions are the following:

Assumption A: The factors in f_t are non-constant variables with finite moments up to the fourth order, $E(f_t) = 0_{L_o \times 1}$ and $E(f_t f_t') = \Omega_f$ for all *t*, and $T^{-1}\Sigma_{t=1}^{T} f_t f_t' \to_p \Omega_f$ as $T \to \infty$, where Ω_f is a $L_o \times L_o$ finite and positive definite matrix.

Assumption B: $rank(\mathrm{B}) = L_o$.

Assumption C: There exists a constant $m \in (0, \infty)$, such that for all *T* (with fixed *N*), (C1) the errors ε_{it} have finite moments up to the eighth order with $E(\varepsilon_{it} \mid f_1, f_2, ..., f_t) = 0$ for all *i* and *t*; (C2) $E(\varepsilon_{it}\varepsilon_{i's} \mid f_1, f_2, ..., f_t) = 0$ for all $i \neq i'$, $s \geq t$; (C3) $\left|T^{-1}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T} E(\varepsilon_{is}\varepsilon_{it})\right| \leq m$ for all *i* and *t*;

(C4) $T^{-1/2}\Sigma_{t=1}^{T} w_t \to_d N\left(0_{(N+L_o)\times 1}, \Lambda\right)$, as $T \to \infty$, where $w_t = (h_t \otimes \varepsilon_{\bullet t}) - E(h_t \otimes \varepsilon_{\bullet t})$, h_t = $(1, f_t', \varepsilon_{\bullet t}')'$, and $\Lambda = p\lim_{T\to\infty} T^{-1}\Sigma_{s=1}^{T}\Sigma_{t=1}^{T} E(w_t w_s')$.

Assumption D: Let B_G be the factor loading matrix corresponding to L ($\geq L_o$) arbitrarily chosen response variables from $r_{\bullet t}$. Then, $rank(\mathrm{B}_G) = L_o$.

In Assumption A, we assume that the factors are covariance stationary; that is, the variance matrix of f_t, $Var(f_t) = \Omega_f$, is same for all t, we adopt this assumption for expository convenience. The assumption can be relaxed without altering our results. The required assumption is that $T^{-1}\Sigma_{t=1}^{T} f_t f_t' \to_p \Omega_f$ as $T \to \infty$. Most of the general mixing processes satisfy this condition (White, 1999).

Assumption B implies that the true number of factors is L_o. Under Assumption (C1), the factors are weakly exogenous to the idiosyncratic errors. Assumption (C2) restricts the error terms to be cross-sectionally uncorrelated. Thus, with (C2), the model (2) is an exact factor model. For the cases in which observable variables correlated with factors are absent, this assumption is crucial for the estimation of the number of factors. Alternatively, when the errors are cross-sectionally correlated, but not autocorrelated over time, an exact model can be obtained by rewriting the model (2) as

$$r_{i\bullet} = F\beta_i + \varepsilon_{i\bullet},$$

where $F = (f_1, f_2, ..., f_T)'$. If the errors are serially uncorrelated, the variance matrix of $\varepsilon_{i\bullet}$ becomes diagonal. When T is small, we can estimate L_o by applying the method we discuss below to this alternative model.

If some instrumental variables correlated with the factors are observable, we could use them to estimate the number of factors, even allowing the errors to be cross-sectionally correlated. Such cases will be discussed in section 4.2.

Assumption (C3) indicates that the autocovariances of the error terms are absolutely summable, while (C4) is nothing but a central limit theorem. When factors and errors follow general mixing processes, both Assumptions (C3) and (C4) hold.

Assumption D implies that all factors (L_o) influence all possible subsets of response variables. In order to motivate Assumption D, let us partition the response variables in $r_{\bullet t}$ into two arbitrary groups.

$$\underset{(N\times 1)}{r_{\bullet t}} = \begin{pmatrix} \underset{(P\times 1)}{g_{\bullet t}} \\ \underset{(Q\times 1)}{z_{\bullet t}} \end{pmatrix} = \underset{(N\times 1)}{\alpha} + \underset{(N\times L_o)}{\mathrm{B}} \underset{(L_o\times 1)}{f_t} + \underset{(N\times 1)}{\varepsilon_{\bullet t}} = \begin{pmatrix} \underset{(P\times 1)}{\alpha^g} + \underset{(P\times L_o)}{\mathrm{B}^g} \underset{(L_o\times 1)}{f_t} + \underset{(P\times 1)}{\varepsilon_{\bullet t}^g} \\ \underset{(Q\times 1)}{\alpha^z} + \underset{(Q\times L_o)}{\mathrm{B}^z} \underset{(L_o\times 1)}{f_t} + \underset{(Q\times 1)}{\varepsilon_{\bullet t}^z} \end{pmatrix}, \tag{3}$$

such that $P+Q=N$, $P>L_o$, and $Q>L_o$. Then, Assumption D, with Assumptions A-C, implies that

$$rank[E(z_{\bullet t}(g_{\bullet t}-\alpha^g)')] = rank[\mathrm{B}^z\Omega_f\mathrm{B}^{g\prime}] = L_o. \tag{4}$$

Based on this observation, we propose to estimate L_o by estimating the rank of $E[z_{\bullet t}(g_{\bullet t}-\alpha^g)']$.

Clearly, Assumption D is stronger than Assumption B. Many of the methods popularly used for factor analysis do not require Assumption D. Under Assumptions A-C,

$$E[(r_{\bullet t}-\alpha)(r_{\bullet t}-\alpha)'] = \mathrm{B}\Omega_f\mathrm{B}' + \Psi,$$

where Ψ is the $N\times N$ diagonal matrix of the variances of ε_{it}. The ML estimation of Jöreskog (1967) and the Minimum Chi-Squared statistic (MINCHI2) of Cragg and Donald (1997) estimate L_o based on estimates of B and Ψ. But use of these methods is somewhat limited. The legitimacy of the ML method requires some strong distributional assumptions on data such as normality. Use of MINCHI2 does not require such strong distributional assumptions, but it often suffers from a computational difficulty in estimating Ψ. Adopting Assumption D, we no longer need to estimate Ψ. It suffices to estimate the rank of the moment matrix $E(z_{\bullet t}(g_{\bullet t}-\alpha^g)')$.

Assumption D requires that most of the response variables should depend on *all* of the factors in f_t. Too see why, suppose that L_o or more response variables in $g_{\bullet t}$ depend on only a subset of f_t; that is, the factor loadings of many (L_o or more) response variables corresponding to a subset of factors are zeros. For such cases, Assumption D is violated depending on the partitions of $g_{\bullet t}$ and $z_{\bullet t}$. We will consider such cases later.

The rank condition (4) can be converted to a moment condition that can be used in GMM. According to Assumption D, there must exist a $P\times(P-L_o)$ matrix $\Xi=(\Xi_1',-\Xi_2')'$ of *full column*, where Ξ_1 is a $(P-L_o)\times(P-L_o)$ square invertible matrix, such that $\mathrm{B}^{g\prime}\Xi = 0_{L_o\times(P-L_o)}$. Thus, under Assumptions A-D, we have

$$E\left[\begin{pmatrix}1\\ z_{\bullet t}\end{pmatrix}(\Xi' g_{\bullet t}-\alpha_{\Xi})'\right]=E\left[\begin{pmatrix}1\\ z_{\bullet t}\end{pmatrix}(g_{\bullet t}-\alpha^{g})'\Xi\right]=E\left[\begin{pmatrix}1\\ z_{\bullet t}\end{pmatrix}\varepsilon_{\bullet t}^{g\prime}\Xi\right]=0_{(Q+1)\times(P-L_o)}, \tag{5}$$

where $\alpha_{\Xi}\equiv\Xi'\alpha^{g}$ is a $(P-L_o)\times 1$ vector. Assumption C(2), which restricts the model (2) to be an exact one, is crucial for this moment condition. For future use, define $\theta = vec[(\alpha_{\Xi},\Xi_2')']$. Clearly, Ξ is not unique, since for any conformable square matrix A, $(\Xi A)'\mathrm{B}^{g}=0$. There are many possible restrictions we can impose to avoid this under-identification problem. Among them, we use the restriction $\Xi_1 = I_{P-L_o}$, while leaving Ξ_2 unrestricted. Among the $P\times(P-L_o)$ matrices satisfying this restriction, Ξ is the unique $P\times(P-L_o)$ matrix of full column that is orthogonal to B^{g}.[3]

3. GMM ESTIMATION OF THE NUMBER OF LATENT FACTORS

In this section we present the GMM method for estimation the number of factors. First, given the assumptions explained before, we construct the moment conditions that will be used in the estimation. Let us denote by L the number of factors we use for estimation, which could be different from L_o. Given L, we partition $g_{\bullet t}$ into

$$\underset{(P\times1)}{g_{\bullet t}}=\begin{pmatrix}\underset{((P-L)\times1)}{y_{\bullet t}^{P-L}}\\ \underset{(L\times1)}{x_{\bullet t}^{L}}\end{pmatrix}=\begin{pmatrix}\underset{((P-L)\times1)}{\alpha_{P-L}^{y}}+\underset{((P-L)\times L_o)}{\mathrm{B}_{P-L}^{y}}\underset{(L_o\times1)}{f_t}+\underset{((P-L)\times1)}{\varepsilon_{P-L,\bullet t}^{y}}\\ \underset{(L\times1)}{\alpha_{L}^{x}}+\underset{(L\times L_o)}{\mathrm{B}_{L}^{x}}\underset{(L_o\times1)}{f_t}+\underset{(L\times1)}{\varepsilon_{L,\bullet t}^{x}}\end{pmatrix}, \tag{6}$$

where $L = 0, 1, 2, \ldots, P-1$. With this notation, define the following moment function:

$$m_t(b_L\,|\,L)=\left[I_{P-L}\otimes\begin{pmatrix}1\\ z_{\bullet t}\end{pmatrix}\right][y_{\bullet t}^{P-L}-(I_{P-L}\otimes(1,x_{\bullet t}'))b_L], \tag{7}$$

where b_L is a $(P-L)(L+1)\times 1$ vector of unknown parameters. Observe that the moment function (7) is linear in b_L. Also note that the moment function (7) is the one implied by a

[3] Specifically, $\Xi_2' = \mathrm{B}_1^{g}(\mathrm{B}_2^{g})^{-1}$ where $\mathrm{B}^{g}=[\mathrm{B}_1^{g\prime},\mathrm{B}_2^{g\prime}]'$, and B_2^{g} is a square invertible matrix.

multiple equation model with $(P-L)$ different dependent variables ($y_{\bullet t}^{P-L}$), with common regressors ($x_{\bullet t}^{L}$) and common instrumental variables ($z_{\bullet t}$). Thus, the moment function (7) can be easily imposed in GMM using any software that can handle three-stage least squares.

The intuition behind moment function (7) comes from the fact that it is linked to moment condition (5). To see why, let $H_L = (I_{P-L}, -S'_{P-L})'$. be a $P \times (P-L)$ matrix with a $L \times (P-L)$ unrestricted parameter matrix S_{P-L}; and let a_{P-L} be a $(P-L)\times 1$ unrestricted parameter vector. By construction, H_L is a full-column matrix. Furthermore, it can be shown:

$$m_t(b_L \mid L) = vec\left[\begin{pmatrix} 1 \\ z_{\bullet t} \end{pmatrix}(H'_L g_{\bullet t} - a_{P-L})'\right]. \quad (8)$$

Thus, the moment condition (5) implies that under Assumptions A-D, when $L = L_o$, $E[m_t(b_L \mid L_o)] = 0$ if and only if $b_L = \theta$. That is, our moment conditions will hold just at the true value of the parameters, and if and only if the true number of factors (L_o) was used in the estimation.

Now, we explain how to use the moment function to consistently estimate the factors. For given L, consider the following minimization problem:

$$\min_{b_L} \; c_T(b_L \mid W_T(L), L) = T d_T(b_L)'[W_T(L)]^{-1}(b_L), \quad (9)$$

where $d_T(b_L \mid L) = T^{-1}\Sigma_{t=1}^{T} m_t(b_L \mid L)$ is the sample mean of the moment functions $m_t(b_L \mid L)$, and the weighting matrix $W_T(L)$ is $(P-L)Q \times (P-L)Q$ positive-definite matrix with a non-stochastic and finite probability limit, say $W(L)$. Let $\hat{b}_L$ denote the GMM estimator minimizing $c_T(b_L \mid W_T(L), L)$; and use $\hat{b}_L^o$ to denote the GMM estimator minimizing $c_T(b_L \mid W_T(L_o), L_o)$ (i.e. at the true number of factors). Let $\tilde{W}_T(L_o)$ be a consistent estimator of $\lim_{T\to\infty} Var\left(\sqrt{T} d_T(\theta \mid L_o)\right)$. The estimator $\tilde{W}_T(L_o)$ can be obtained by using the method of White (1980) if data are serially uncorrelated, and the methods of Newey and West (1987) or Andrews (1991) if data are serially correlated. We now denote by $\tilde{b}_L^o$ the optimal GMM estimator of θ that minimizes $c_T(b_L \mid \tilde{W}_T(L_o), L_o)$. Using this notation, the following result establishes that the moment conditions on (7) can be used to estimate the number of factors.

Proposition 1: Under Assumptions A-D, for any $W_T(L)$, $c_T(\hat{b}_L \mid W_T(L), L) \to_p \infty$ for any $L < L_o$ and $c_T(\hat{b}_L^o \mid W_T(L_o), L_o) \to_d \Upsilon$, where Υ is a weighted average of independent $\chi^2(1)$ random variables. In addition, $c_T(\tilde{b}_L^o \mid \tilde{W}_T(L_o), L_o) \to_d \chi^2[(P - L_o)(Q - L_o)]$.

The proof of Proposition 1 is given in the appendix. The distribution of $c_T(\hat{b}_L \mid W_T(L), L)$ is generally unknown, but the results from Proposition 1 are sufficient to derive the estimation methods for the number of factors. We can formulate the model (2) assuming different number of factors (i.e. different values of L) and then, use the c_T statistics to select the model that has the best fit. Two approaches have been proposed in the literature for model selection. The first one uses a sequential hypothesis testing approach, and the second is based on model selection criterion. We can apply these two approaches to estimate the number of factors.

Our sequential testing approach is based in the asymptotic distribution of $c_T(\tilde{b}_L^o \mid \tilde{W}_T(L_o), L_o)$ statistic, which is simply the overidentifying restriction test statistic (Hansen, 1982). Using this approach, we first formulate the factor model (2) assuming that the true number of factors is equal to one (L_o = 1). Then we estimate b_L by GMM, compute the overidentifying restriction statistic, and test the hypothesis of L_o = 1 against the alternative hypothesis of L_o > 1. By proposition 1, if L_o is greater than one, the statistic diverges to infinity in large sample. Thus, we can expect that the test is likely to reject the hypothesis of $L_o = 1$, if the sample size is reasonably large. If the hypothesis is rejected, we will formulate the model (2) with L = 2, and compute the overidentifying restriction statistic to test the null hypothesis of L_o = 2 against the alternative of L_o > 2. We continue this procedure until the null hypothesis is not rejected. This sequential procedure can yield a consistent estimator of L_0 if an appropriate adjustment is made to the significance level used for the test. The adjustment is necessary because type 1 errors are accumulated as the test continues. Cragg and Donald (1997) show that the significance level α_T should be adjusted such that $\alpha_T \to 0$, and $-\log \alpha_T / T \to 0$ as $T \to \infty$.

The model section criterion method has been used extensively in determining the order of ARMA processes in time series analysis, specifically by Hannan and Quinn (1979), Hannan (1980,1981), Atkinson (1981), and Nishii (1988). Cragg and Donald (1997) use this method to estimate the ranks of matrices. Following these studies, we define the following criterion function:

$$MS_T(L) = c_T(\hat{b}_L \mid W_T(L), L) f(T)^{-1} - g(L), \tag{10}$$

where $f(T)$ and $g(L)$ are predefined functions of T (the number of observations) and L (the number of factors), respectively. With appropriate choices of $f(T)$ and $g(L)$, a consistent estimate of L can be obtained by minimizing the criterion function $MS_T(L)$. There are many possible choices of $f(T)$ and $g(L)$. One commonly used criterion is:

Schwarz Criterion (BIC): $f(T) = \ln(T)$, and $g(T) = (P-L)(Q-L)$.

In BIC, $g(L)$ is simply the degrees of overidentifying restrictions in the moment condition $E[m_t(b_L \mid L)] = 0$. With (10) and BIC, we obtain the following result:

Proposition 2: Let $\hat{L}$ be the minimizer of $MS_T(L)$ with BIC. Then, $\hat{L} \to_p L_o$.

The proof of Proposition 2 is given in the appendix, even though it is a straightforward extension of a result from Ahn, Lee and Schmidt (2007b). They have studied a panel data model with latent components of factor structure. They developed a GMM method to estimate the model and the number of factors in the latent components with BIC. Their results are easily extended to our factor model. Interested readers may refer to the paper.

Observe that Proposition 2 holds even if the optimal GMM estimator is not used. One important advantage of the criterion method over the sequential method is that it does not require use of the optimal GMM estimator. In the GMM literature, many studies have shown that optimal GMM estimators often have poor finite-sample properties, especially when data are autocorrelated or/and too many moment functions are used (see, for example, Altongi and Segal, 1996; Andersen and Sørensen, 1996; and Christiano and den Haan, 1996). One of the main reasons for this problem is that for such cases, the optimal weighting matrix, $[\tilde{W}_T(L_o)]^{-1}$ is poorly estimated. Given this problem, in practice, the selection criterion method appears to be an attractive alternative to the sequential method.

The sequential testing and model selection criterion methods can consistently estimate L_o if Assumption D holds. However, as we have discussed above, the assumption would be violated if some factors influence only a subset of the response variables. When the assumption does not hold, our methods tend to underestimate the number of factors. To see why, consider the following alternative assumption:

Assumption D*: $rank(\mathrm{B}^z\Omega_f\mathrm{B}^{g\prime}) = L^* \le L_o$, and $rank[\mathrm{B}^z\Omega_f(\mathrm{B}^x_{L^*})'] = L^*$.

In the appendix (Lemma A.1), it is shown that when $L = L^*$, a unique vector θ^* exists such that $E[m_t(\theta^* \mid L^*)] = 0$. Let $\tilde{W}_T(L^*)$ be a consistent estimator of $\lim_{T\to\infty} Var\left(\sqrt{T}d_T(\theta^* \mid L^*)\right)$; and let $\hat{b}_L^*$ and $\tilde{b}_L^*$ be the minimizers of $c_T(b_L \mid W_T(L^*), L^*)$ and $c_T(b_L \mid \tilde{W}_T(L^*), L^*)$, respectively. Then, by replacing Assumption D by D*, we obtain the following results:

Proposition 3: Under Assumptions A-C and D*, for any choice of $L < L^*$ and $W_T(L)$, $c_T(\hat{b}_L \mid W_T(L), L) \to_p \infty$. In contrast, $c_T(\hat{b}_L^* \mid W_T(L^*), L^*) \to_d \Upsilon$, where Υ is a weighted average of independent $\chi^2(1)$ random variables. In addition, $c_T(\tilde{b}_L^* \mid \tilde{W}_T(L^*), L^*) \to_d \chi^2[(P-L^*)(Q-L^*)]$.

Since the partition of $g_{\bullet t}$ and $z_{\bullet t}$ is arbitrary, the rank of $\mathrm{B}^z\Omega_f\mathrm{B}^{g\prime}$ could change depending on the choice of $g_{\bullet t}$ and $z_{\bullet t}$ if Assumption D does not hold. Thus, Proposition 3 indicates that when Assumption D is violated, the estimated number of factors could be sensitive to the partition used in estimation. As a treatment to this problem, we propose to try many different partitions to estimate the number of factors. We can try a subset of all possible partitions, or some randomly generated partitions. Our simulation exercises show that using the frequency table of the estimates from a sufficiently large number of different partitions, we can obtain an accurate estimate the correct number of factors.

4. Extensions

In this section, we consider the two cases to which the GMM methodology developed in the previous section can be generalized.

4.1. Approximate Factor Models

Chamberlain and Rothschild (1983) propose an *approximate factor model* to test the Arbitrage Price Theory. This model differs from the exact factor model since it allows idiosyncratic components to be cross-sectionally correlated. Assumption C implies that $Var(\varepsilon_{\bullet t}) \equiv \Psi$ is diagonal. In contrast, the approximate factor model allows Ψ to be non-diagonal, although the correlations among the errors in $\varepsilon_{\bullet t}$ are restricted to be mild. Chamberlain and Rothschild (1983) have shown that for an approximate model with L_o factors, the first L_o eigenvalues of the variance matrix of the response variables diverge to infinity as $N \to \infty$, while other eigenvalues remain bounded. Based on this finding, they suggest estimating L_o by counting the number of larger eigenvalues of the variance matrix of response variables. Bai and Ng (2002) proposes a more elaborated statistical method. These two methods are appropriate for the data with both large N and T. However, they may not be appropriate for the data with small N (see Brown, 1989; Bai and Ng, 2002).

While our method is designed for exact factor models with small N, it could be used to estimate some approximate factor models. For example, consider a model in which the response variables in $r_{\bullet t}$ are categorized into a finite number (M) of groups (e.g., portfolios). Each of the groups, indexed by G_1, G_2, … , G_M, contains NG_j variables, such that $\Sigma_{j=1}^{M} NG_i = N$, and for all $j = 1,...,M$, $NG_j / N \to a_j$ for some positive number a_j, as $N \to \infty$. Suppose that the response variables are generated by the following processes:

$$r_{j,it} = \alpha_{j,i} + (\beta_{j,i}^{glo})' f_t^{g} + (\beta_{j,i}^{loc})' f_{j,t}^{l} + u_{j,it}, \tag{11}$$

where i indexes individuals, $j = 1,...,M$ indexes individual groups, the variables in f_t^{glo} are the "global" factors that influence all of the response variables in different groups, the variables in $f_{j,t}^{loc}$ are the "local" factors that are correlated with the variables in group j, but not with those in other groups (e.g., $E(f_{j,t}^{loc} f_{j',t}^{loc}) = 0$, for $j \neq j'$), the $\alpha_{j,i}$ are intercept terms, and the vectors $\beta_{j,i}^{glo}$

and $\beta_{j,i}^{loc}$ are the loadings of the corresponding factors. The $u_{j,it}$ are idiosyncratic errors. Approximate factor models restrict the cross-section correlations in the error terms to be mild. For example, Bai and Ng (2002) impose the following restriction, which we name "Approximate Assumption" (AA):

Assumption AA: Let $\tau_{ii',ts} = E(u_{it}u_{i's})$, where u_{it} and $u_{i't}$ are the error terms from the same or different groups, and t and s are time indexes. Then, $(NT)^{-1}\Sigma_{t=1}^{T}\Sigma \ \left|\tau_{ii',tt}\right| \le \left|\tau_{ii'}\right|$ for some $\tau_{ii'}$ and for all t, and $N^{-1}\Sigma_{i=1}^{N}\Sigma_{i'=1}^{N}\left|\tau_{ii'}\right| \le M$ for some positive number M, for all N.

Let $\overline{u}_{j,t} = (NG_j)^{-1}\Sigma_{i\in G_j}u_{j,it}$. Then, Assumption AA warrants that N, $T \to \infty$,

$$E(\overline{u}_{j,t}\overline{u}_{j',t}) = \frac{1}{NG_j \times NG_{j'}}\Sigma_{i\in G_j}\Sigma_{i'\in G_{j'}}u_{j,it}u_{j',i't} \to 0; \tag{12-1}$$

$$\frac{1}{T}\Sigma_{t=1}^{T}\overline{u}_{j,t}\overline{u}_{j',t} = \frac{1}{T \times NG_j \times NG_{j'}}\Sigma_{t=1}^{T}\Sigma_{i\in G_j}\Sigma_{i'\in G_{j'}}u_{j,it}u_{j',i't} \to_p 0. \tag{12-2}$$

Now, consider the following group-mean equations of (11):

$$\overline{r}_{j,t} = \overline{\alpha}_j + (\overline{\beta}_j^{glo})f_t^g + (\overline{\beta}_j^{loc})' f_{j,t}^{loc} + \overline{u}_{j,t} = \overline{\alpha}_j + (\overline{\beta}_j^{glo})f_t^g + \overline{\varepsilon}_{j,t}, \tag{13}$$

where the symbols with overhead bar are defined similarly to $\overline{u}_{j,t}$. By (12-1)-(12-2) and the fact that the variables in $f_{j,t}^{loc}$ are group-specific, we can show that $\overline{\varepsilon}_{j,t}$ are asymptotically uncorrelated over different groups. That is, we can treat the equations in (13) as an exact factor model if N and NG_j are sufficiently large. Thus, using our method, we could estimate the number of the global factors by estimating the rank of $\mathrm{B}^{glo} = \lim_{N\to\infty}\left(\overline{\beta}_1^{glo}, \overline{\beta}_2^{glo}, ..., \overline{\beta}_n^{glo}\right)'$.

4.2. GMM Estimation with Observable Instruments

When some variables are potentially correlated with latent factors and observable, we could use them to estimate L_o, or test how many of them are indeed correlated with the factors. We first consider how to estimate L_o. Let s_t be the $K \times 1$ vector of instruments which satisfies following assumption:

Assumption D**: $rank[E(s_t f_t')] = L_o < K$ and $E(\varepsilon_{\bullet t} s_t') = 0_{N\times K}$.

Under Assumption D**, there must be a $N \times (N - L_o)$ matrix of full column, Ξ^{**}, such that

$$E\left[\begin{pmatrix} 1 \\ s_t \end{pmatrix}(r_{\bullet t} - \alpha)'\Xi^{**}\right] = 0_{(K+1)\times(N-L_o)}. \qquad (14)$$

Thus, we can estimate L_o using the same method discussed in section 3. Our methods apply as we use $r_{\bullet t}$ for $g_{\bullet t}$ and s_t for $z_{\bullet t}$. When observable instruments are not available, we need to partition response variables into two groups to use a group of response variables as the instruments for latent factors. But for the response variables in a group to be legitimate instruments, the error terms in $\varepsilon_{\bullet t}$ should be cross-sectionally uncorrelated. When outside instruments are observable, we do not need to partition the response variables. In additions, the error terms are allowed to be cross-sectionally correlated as long as the instruments are not correlated with them.

In cases in which the number of factors is already known, or estimated by the methods discussed in section 3, we can test by GMM how many of the factors are correlated with the observable instrumental variables in s_t. If some factors are not correlated with s_t, it should be the case that $rank[E(s_t f_t')] = L^{**} < L_o$. For this case, by the same method we used in section 3, we can show that the GMM methods based on the moment condition (14) estimate L^{**}, not L_o.

5. MONTE CARLO SIMULATION

5.1. Data Generation

The foundation of our Monte Carlo exercises is the following the three-factor model:

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \beta_{i3} f_{3t} + \varepsilon_{it} = \alpha_i + c_{1,it} + c_{2,it} + c_{3,it} + \varepsilon_{it}, \qquad (15)$$

where the f_{kt} (k = 1, 2, 3) are the common factors of the model. Our benchmark model is the three-factor model of Fama and French (1993): EMR (excess market return), SMB, and HML.[4]

[4] The Fama-French factors are constructed using the 6 value-weight portfolios formed on size and book-to-market. SMB (Small Minus Big) is the average return on the three small portfolios minus the average return on the three big portfolios. HML (High Minus Low) is the average return on the two value portfolios minus the average return on

We generate randomly β_{ik} and f_{kt} to match the moments of the Fama-French data. That is, we generate data such that the moments of $c_{k,it}$ match the counterparts from the data that Fama and French (1993) used. At the sample means of the estimated betas ($\bar{\beta}_1$, $\bar{\beta}_2$, and $\bar{\beta}_3$) for the 25 size and book-to-market portfolios, the estimated variances of the Fama-French common components are the following:

$$var(\bar{\beta}_1 \times EMR) = 21.72; var(\bar{\beta}_2 \times SMB) = 4.50; var(\bar{\beta}_3 \times HML) = 1.29.$$

Two types of idiosyncratic error components are used. First, we generate the errors which are cross-sectionally heteroskedastic, but not autocorrelated. Specifically, the errors are drawn from $N(0, \sigma_i^{FF})$, where the σ_i^{FF} are the variances of the residuals from the time-series regressions of (15) for each i. The values of σ_i^{FF} are between 1.21 and 3.78, with the average of 2.016. Thus, the variances of the first and second common components at the means of betas are more than twice as great as the average variance of the idiosyncratic components, while the variance of the third common component (1.29) is smaller. We define the signal to noise ratio (SNR) of a common component ($c_{k,it}$) as the ratio of the variances of the common component and the idiosyncratic error component. In our simulation, the SNRs are 10.8 , 2.2, and 0.65 for common components 1, 2, and 3, respectively.

Second, we generate the error terms from a simple AR (1) process: $\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + v_{it}$. Using the residuals from the time-series regressions of (15), we estimate the parameters ρ_i and estimate $\mathrm{var}(v_{it})$ such that $\mathrm{var}(\varepsilon_{it}) = \sigma_i^{FF}$. The errors generated by this way are cross-sectional heteroskedastic and serially correlated over time.

5.2. Size of the Over-Identifying Restriction Tests

We first investigate the size of the overidentifying restrictions test with the true number of factors. We use *N* portfolio returns generated by equation (15). We randomly divide the *N* portfolios into two groups: $g_{\bullet t}$ and $z_{\bullet t}$ in the notation of section 2. Then, we carry out the

---

the two growth portfolios. EMR is the excess return on the market: the value-weight returns on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates). See Fama and French (1993) for a complete description of the factor returns.

GMM estimation discussed in section 2,[5] compute the overidentifying restriction test statistic (which we simply call "J statistic" from now on), and finally, test if the true number of factors is equal to 3, at a significance level of 5%. We proceed to repeat this procedure for 1,000 iterations, and compute how many times the true null hypothesis ($L_o = 3$) is rejected. Since the portfolio returns are generated by 3 factors we expect to reject the null hypothesis 5% of the times.

We perform our simulations with six different combinations of T and N: T = 500 and 1000; N = 12, 15, and 25. For each combination, we consider two cases: the cases with autocorrelated (AR(1)) and serially uncorrelated idiosyncratic errors. We also try different numbers of instruments. That is, we conduct simulations using different number of portfolios in each of the two groups ($g_{\bullet t}$ and $z_{\bullet t}$) of the partition of total N portfolios. For the cases with N = 12, we try three different partitions: (P,Q) = (8,4), (7,5) and (6,6). We also try (P,Q) = (9,6), (7,8) and (6,9) for N = 15; and (P,Q) = (17,8), (16,9), and (13,12) for N = 25. We perform this experiment to check whether the test results are sensitive to the number of instruments (Q) used. There have been many studies finding that the GMM estimators computed with too many instruments and small data are often biased (see, for example, Andersen and Sørensen, 1996).[6] The values of N and T are chosen to be close to the sample sizes most often used in the finance literature. The percentage of rejection of the true null hypothesis by the J test statistic is presented in Table 1.

For the case of no autocorrelation, the J test performs relatively well for all of the specifications we experimented. It appears that the test performs better when 5, 8, and 9 instruments (Q) are used for the data with N = 12, 15, and 25, respectively (not counting the vector of ones as instrument). We suggest that in practice initially the partition should include around half of the response variables (N/2) as instruments, but the number of instruments (Q) should not be greater than 10.

As expected, the size of the test improves as the number (T) of times series observations increases from 500 to 1000. For the cases of autocorrelated errors, we use the Newey-West

[5] We develop a programs using GAUSS 6.0 for the estimation and data generation.

[6] Using only a small subset of the available moment conditions is not a solution either. Andersen and Sørensen (1996) showed that using too few moment conditions are as bad as estimators using too many conditions. This result indicates that there is a trade-off between informational gain and finite-sample bias caused by using more moment conditions.

(1987) covariance matrix to compute the optimal weighting matrix. We present the results obtained using two different bandwidths: 3 and 0.[7] Notice that, if bandwidth = 0, the Newey-West matrix reduces the heteroskedasticity-robust variance of White (1980).

When the idiosyncratic errors are autocorrelated, the White variance matrix is not the optimal choice. Our results indeed show that when autocorrelation is present in data, the test results are quite sensitive to the choices of bandwidth and number of instruments, especially for the data with large N. For N = 12, the test performs better when the statistic is computed with bandwidth = 3. When N = 15 or 25, and bandwidth = 3, the test statistic under-rejects the true null hypothesis of $L_o = 3$ for almost any choice of the number of instruments. When bandwidth = 0 (which is asymptotically not an optimal choice) is selected for the cases with N = 15 and 25, the test is rather better sized. It appears that the Newey-West estimator becomes less reliable when N is large. It may be so because the number of parameters in the weighting matrix rapidly increases with N. For the cases of N = 15 and 25, the efficiency gain by using the estimated optimal weighting matrix do not seem to be large enough to compensate for the loss by using poorly estimated weighting matrix.

5.3. Estimation of the number of factors

Using the data generated using three factors as defined in section 5.1, we now estimate the number of factors using the sequential hypothesis testing and the model selection criterion methods. As discussed in section 3, to obtain consistent estimates by using the sequential test method, we need to adjust the significance level (α_T) depending on the sample size (T). We use $\alpha_T = 0.05 \times \sqrt{500/T}$. This function is chosen such that $\alpha_{500} = 0.05$.[8] 1,000 different sets of randomly generated portfolio returns are used for simulations. The results are summarized in Tables 2 and 3.

Table 2 shows that for all of the different combinations of N and T, the sequential hypothesis testing method produces quite reliable estimates when the idiosyncratic errors are not

[7] In unreported experiments we use other different choices of bandwidth. Results and main conclusions do not change. The automatic bandwidth selection methods by Andrews (1991) or Newey and West (1994) chose the values of bandwidth greater than six for our simulated data, but with the values greater than six, our test results get worse. The tests with bandwidth of three performed better.

[8] In unreported experiments, we also have tried many other significance levels, but the results do not show remarkable changes.

autocorrelated and the heteroskedasticity-robust weighting matrix is used in GMM. As expected, the estimates become more accurate as T increases. We obtain less accurate results when $N = 500$ and $T = 25$: The estimated numbers of factors are three for 93.20% of the times. For the cases of autocorrelated errors, we obtain the similar results even if the heteroskedasticity-robust weighting matrix (which is not optimal) is used. We obtain the similar results using the Newey-West matrix with bandwidth = 3, except for one case. For the simulated data with $N = 25$ and $T = 500$, the sequential test with bandwidth = 3 predicts one factor for 100% of the times. This result is consistent with the results of size in Table 1, which shows that the J test tends to under-reject the true null hypothesis. Our simulations results from the sequential hypothesis testing method suggest that larger samples are required to analyze a large number of portfolios ($N \geq 15$) when idiosyncratic errors are autocorrelated.

The results reported in Table 3 show that the model selection criterion method is slightly better than the sequential test method in estimating the number of factors, for almost all of the different combinations of T and N. Similarly to the results from the sequential tests for the data with $N = 25$, and $T = 500$, the numbers of factors estimated by the selection criterion method are severely downward biased when the bandwidth of three is used. As we find out in section 3, the model selection criterion procedure does not require use of the optimal weighting matrix. Thus, the results reported in Tables 2 and 3 suggest that the model selection criterion method using the heteroskedasticity-robust variance matrix performs well whether or not the idiosyncratic errors are autocorrelated.

To check the robustness of our results, we now investigate if the estimation results change substantially when we change our partitions of portfolio returns into two groups. To accomplish this objective we generate one set of portfolio returns, and then, we randomly create 100 different partitions. Then, for each partition, we estimate the number of factors. The results from the experiments with $T = 1{,}000$ are presented in Table 4 Panel A.[9] The sequential testing method estimates the number of factors more accurately when $N = 12$ or 15, than when $N = 25$. We find the correct number of factors more than 89% of the times when $N = 12$ or 15, but around 65% when $N = 25$. The model selection criterion method produces more reliable estimates,

[9] We do not report results for T=500 in order to save space and since the main conclusion of this section do not are not different. Results are available upon request.

especially for the cases with N = 15 and 25. We obtain the correct number of factors over 91% when N = 15 and over 82% when N = 25.

This experiment confirms that the GMM estimation results could change depending on the partition we use. The test results are more sensitive to the partitions used in GMM when N is large. The model selection criterion method is less sensitive to the partitions than the sequential testing method does. Nonetheless, the selection method is also likely to produce incorrect inferences when the data with large N are analyzed. For this reason, we propose to estimate the number of factors using many randomly partitioned data. Our experiments suggest that the number of factors can be more accurately estimated if 100 different partitions are used for estimation.[10] The number of factors most often estimated from different partitions could be a reliable estimator (i.e. the one with the highest frequency). In order to confirm this conjecture, we perform the following experiment: We generate 1,000 different sets of portfolio returns. For each data set, we estimate the number of factors using 100 randomly created partitions and choose as our point estimator the number estimated most often. The results are presented in Table 4. Panel B of Table 4 confirms our conjecture. The number of factors most often estimated from different partitions is always 3, for N = 12, 15, and 25, whether or not idiosyncratic errors are autocorrelated or not.

The last part of our simulations tries to evaluate the performance of our methods when a factor explains a very small proportion of the total variation of the response variable. We will call such factor a *weak* factor. In our simulation, the variance of the common component ($c_{k,it} = \beta_{ik} f_{kt}$) associated to a weak factor will be small compared with the variance of the idiosyncratic component. In order words, a weak factor is a factor with a low signal to noise ratio (SNR). As described in section 5.1, our data was generated using as a benchmark the three-factor model of Fama and French (1992), where the SNRs of three factors are 10.8, 2.2, and 0.65, respectively. To generate the data with one weak factor and two other non-weak factors, we reduce the SNR of the second common component (SMB) and increase the one of the first common component (EMR). We do so because we wish to generate data such that the total variations in the response variables explained by the three factors and the variations in

[10] We also performed the same experiment using all possible partitions of portfolio returns into two groups. Results do differ significantly with the ones presented just using 100 random specifications of the groups.

idiosyncratic errors remain constant. In this experiment, we reduce the SNR of the second common component to four different values: 1.0, 0.50, 0.35, and 0.25. As we have done in the before, we generate one set of portfolio returns, and then, we randomly create 100 different partitions. For each partition, we estimate the number of factors. The results are presented in Table 5. Since the model selection criterion method appears to be superior to the sequential method, we only report the results from the former method.

When the SNR of the second common component is greater than or equal to 0.50 (Panels A and B), the model selection criterion method estimates most repeatedly three factors for all of the different values of N. For example, when SNR = 1, we estimate three factors 91% of the time for the data with N = 12, 90% for N = 15, and 75% for N = 25. The second highest frequencies are smaller that 25% in all the cases. These results are very similar in magnitude with the ones presented in Panel A of Table 4 in which the SNR of the second common component is set at 2.2.

When the SNR drops to 0.35, the method also estimates most repeatedly three factors, but with lower frequencies. For example, for the cases with N = 12, 15 and 25, we estimate three factors, 60%, 67%, and 59% of the time, respectively. For N = 12, two factors are estimated 35% of the time, which is the second highest frequency. All the other frequencies are smaller than 25%. Finally, when the explanatory power of the second factor becomes even weaker, that is, when its SNR drops further to 0.25, we estimate most repeatedly two factors. For the cases with N = 12, 15 and 25, we estimate two factors 75%, 60% and 68% of the time, respectively.

While it is somewhat arbitrary, given these results, we define the factor with SNR equal to 0.25 as a "weak" factor. Our simulation results show that when a weak factor is present in data, our estimator (the one that receives the highest frequency from the estimation with 100 random partitions) underestimates the true number of factors. That is our estimator is able to detect factors with a SNR bigger than 0.25. However, the true number of factors (L_o = 3) is estimated with the second highest frequency at the range of 25-30%. Our results suggest that if the second highest frequency in the estimation is larger that 25%, then one of the factors is weak and the true number of factors can be underestimated.

The number of factors most often estimated from 100 different partitions (i.e. the one with the highest frequency) appears to be a reliable estimator unless weak factors are present. In order to confirm this fact, we carry out the same experiment we conduct before: we generate one set of portfolio returns, and then, we randomly create 100 different partitions. For each partition, we estimate the number of factors. Then, our point estimator is the number most often estimated. We repeat this experiment for 1,000 generated samples with four different SNRs of the second common component: 1.0, 0.5, 0.35 and 0.25. The results are presented in Table 6. When SNR is greater than or equal to 0.50, the method estimates 3 factors 100% of the time for all values of *N*. This implies that if all factors are strong, we will have a very reliable estimator of the true number of factors. As SNR gets smaller our method estimates three factors with less often and two factors with more frequency. For example if SNR = 0.35 and $N = 12$, three factors are estimated with the highest frequency 61.70%, but we also estimate two factors 38.30% of the time. When $N = 12$ and SNR = 0.25, we will estimate two factors 89.10% of the time, and the three factors just 10.90% of the time. An important fact from these simulations is that the method never overestimates the number of factors, even in the presence of a weak factor.

We confirm that if all factors are strong (SNR larger than 0.25) the number of factors most often estimated from different partitions is a very accurate estimator of the true number of factors (i.e. the one with the highest frequency). If a factor is weak, this method can lead to underestimation of the number of factors. A second highest frequency larger that 25% will be evidence for the presence of a weak factor.

Based on this conclusions, we suggest researchers should estimate the number of factors with different specifications of two groups (100 randomly generated partitions appear to be enough). The relative frequencies of estimated numbers of factors should be observed. If there is just one frequency larger than 25%, the correct number of factors is the one with this higher frequency. One could judge that no weak factor is present. A second highest frequency larger than 25% could be viewed as an evidence of a weak factor. In this case just looking at the highest larger frequency may lead to underestimate of the number of factors. If the second highest frequency ($\geq 25\%$) corresponds to a number of factors that is greater than the one with the highest larger frequency, then the correct number of factors will be the one with the second largest frequency.

6. EMPIRICAL RESULTS

As an empirical application we estimate the number of factors that explain the stock returns in the United States following an APT model. In order to check the sensibility of the estimation to the number of portfolios used, we employ return data of 10, 15, and 25 portfolios. For the case of the 10 portfolios we use the 10 momentum portfolios obtained from Kenneth R. French's on-line data library. These portfolios are formed as the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios.[11] Under this portfolio design the number of assets in each portfolio is roughly the same for every time period. As detailed in section 4.1, this is important for our estimation since the methodology will perform better if the number of assets in each portfolio is large.[12] The 15 and 25 portfolios are constructed randomly, in the sense that the asset returns included in each portfolio are randomly selected. Again this portfolio design warranties that the number of assets included in each portfolio is roughly the same, so the number of assets in each portfolio is large every time period. Our time series sample includes 954 monthly observations for each portfolio from 1927 to 2006.

We use the model selection criterion since according to simulations it produces more reliable inferences in finite samples than the sequential testing method. We repeat the estimation for 100 randomly selected partitions. Instrument groups include 5, 8 and 9 instruments for N = 10, 15, 25 respectively. The results are very robust to changes in the number of instruments. We use the heteroskedasticity-robust variance matrix for the GMM weighting matrix. Estimation results are presented in Table 7 panel A.

In the case of the 10 momentum portfolios, we estimate 71% of the time three factors. Since this is the only frequency larger than 25% we can conclude that the number of factors in this case is three, and there is not evidence of a weak factor. For the data with 15 portfolios we estimate three factors 50% of the time and 35% two factors. Again three factors are estimated with the highest frequency, but in this case two factors are obtained more than 25% of the time which evidences that one of the three factors is weak. Finally for the 25 portfolios, two factors

[11] More details of portfolio formation can be obtained from Kenneth R. French's on line data library.

[12] The idiosyncratic components of individual stock returns may be cross-sectionally correlated. However, as discussed in section 4.1, the errors of portfolio returns are less likely to be cross-sectionally correlated if each portfolio is constructed with many individual stocks. We use portfolios that maximize the number of assets in each portfolio for every period of time.

are estimated 63% of the time and the second highest frequency corresponds to three factors which are estimated 25% of the time. These results imply that three factors explain the stock market returns, and there is evidence that one of the factors is weak in the sense that its corresponding SNR is smaller than 0.25.

We want to investigate further the 25 portfolios case to verify the presence of a third weak factor. Assumption D in our model implies that all portfolios should be affected by the same number of factors. If just and small number of the portfolios are affected by a third factor then this third factor will be weak. In order to investigate how different groups of portfolios from the total 25 are affected by a third weak factor, we estimate the number of factors for different groups or subsets from the 25 portfolios. Specifically we randomly select 10 portfolios from the 25 portfolios. Results of the estimation of the number of factors for 2 different random selections of 10 portfolios are presented in Table 7 panel A. We estimate two factors with the highest frequency in all the cases. We estimate three factors with frequencies that range from 19% to 31%. This implies that the third factor affects a small group of 25 the portfolios, which confirms the fact that the third factor is weak.

We conclude that 3 factors explain the returns of the US stock market. One of the three factors is a weak factor in the sense that its contribution to the total variance of the stock returns is small relative to the idiosyncratic variance.

In a second part of our empirical exercise we want to investigate which variables proposed in the financial literature as factors are really correlated with the three latent factors. As explained in section 4.2, our estimation methodology can be applied for this purpose by just using the observable variables as instruments in the moment conditions. We will focus in the variables proposed by Chen Roll and Ross 1986 (CRR) and Fama French 1996 (FF). CRR select as factors variables that affect the discount rate used to discount future expected cash flows and variables that influence the expected cash flows themselves. The proposed factors are term structure (UTS), changes in expected inflation (DEI), unexpected inflation (UI) , industrial production growth (MP) and the risk premium (URP). UTS is calculated by subtracting the 10 year constant maturity US government bond and the 4 week Treasury bond yield[13]. DEI is obtained by calculating the monthly changes in expected inflation obtained from the U. of

[13] Data obtained fro the Federal Reserve statistical release H15.

Michigan consumer expectations survey.[14] UI is the difference between the realized inflation for month *t* and the expected inflation for month *t*.[15] MP is the monthly growth in the industrial production index[16]. URP is calculated by the difference between the yield of the US government bond and the Moody's seasoned Baa corporate bond yield. Fama and French three factors: excess market return, SMB, and HML are factor mimicking portfolios formed using firm characteristics and returns as explained in section 5.1. We focus in 342 observations stating in 1978, motivated in the availability of the all data for this sample period.

The first step of our estimation involves estimating the number of factors for this sub-sample of 342 periods. Our simulations showed that in small samples it is better to use small number portfolios, so we use 10 portfolios formed on Momentum. Results are presented in table 7 panel B. Interestingly we estimate 3 factors 78% of the times, similarly that the full sample case. Given that the number of latent factors is three, we proceed to test how many of them are correlated with the observable variables of CRR and FF. Results are presented in Table 7 panel B. First we test for the case CRR non inflation variables (MP, UTS and URP) and we find that using them as instruments we are just able to estimate one factor. This implies that these three variables are able to capture (or are correlated) with just one of the three latent factors. In a second estimation we include all CRR variables. Results do not improve, since just one factor is captured again. Finally we include all CRR variables and the three FF factors in our estimation. Results show that just two unobservable factors are correlated with the set of 8 observable variables since two factors are estimated 78% of the times. We conclude the set of variables included in CRR are correlated with one of the three latent factors and the FF factors are able to capture an extra latent factor. One of the three unobservable factors is not captured by the variables in CRR and FF.

7. CONCLUDING REMARKS

In this paper we present a method to consistently estimate the number of factors in a linear factor model. The test is independent of the factors, since it is assumed that they are unobservable. Since GMM is used it is feasible to allow for time series autocorrelation and heteroskedasticity

[14] Data obtained from the Federal Reserve Economic Data FRED.
[15] The Consumer Price Index obtained from FRED was used to calculate monthly inflation.
[16] The industrial production index used to calculate MP was obtained from FRED

and cross sectional heteroskedasticity of disturbances. Monte Carlo simulations show that the size of the test is better when N is small and T is large.

We have considered two procedures to estimate the number of factors: sequential testing and model selection criterion methods. Our simulations show that the model selection criterion method is more precise, especially for 15 or more response variables. We recommend use of the model selection method since it does not require bandwidth selection or adjustment of significance levels.

Since the method requires the partition of the response variable in two groups, we recommend to estimate the number of factors for different specifications of two groups (100 randomly generated partitions appear to be enough). Simulation show that the number of factors most often estimated from different partitions is a very reliable estimator (i.e. the one with the highest frequency). A second highest frequency larger than 25%, will evidence the presence of a weak factor, i.e. a factors that explains small proportion of the variation of the response variable (SNR smaller than 0.25). In this case just looking at the highest larger frequency may lead to underestimate of the number of factors.

Our empirical results imply that the US stock returns are determined by three factors. There is also evidence that one of the factors is a weak factor. Also we find that the variables proposed by Chen Roll and Ross 1986 are able to capture just one of the three latent factors. Fama and French 1996 are able to capture an extra latent factor. One of the three unobservable factors is not captured by the factors proposed by Chen Roll and Ross 1986 and Fama and French 1996.

APPENDIX

The following lemma is useful to prove Propositions 1 and 3.

Lemma 1: Suppose that the factor model (2) satisfies Assumptions A-C and D*. Then, (i) for $L < L^*$, there exist no values of b_L such that $E[m_t(b_L \mid L)] = 0$. (ii) For $L = L^*$, a unique solution $b_L = \theta^*$ exists for $E[m_t(b_L \mid L^*)] = 0$. Finally, (iii) for $L > L^*$, there are infinitely many b_L's such that $E[m_t(b_L \mid L)] = 0$.

Proof: By construction, $rank(H_L) = P - L$. But, under Assumption D*, the maximum number of the linearly independent columns of H_L that are orthogonal to all of the columns of $\mathrm{B}^z \Omega_f \mathrm{B}^{g\prime}$ is $(P - L^*)$, which is smaller than $(P - L)$. This means that some columns of H_L cannot be orthogonal to all of the columns of $\mathrm{B}^z \Omega_f \mathrm{B}^{g\prime}$ when $L < L^*$. That is, $\mathrm{B}^z \Omega_f \mathrm{B}^{g\prime} H_L \neq 0$. Under Assumptions A-C and D*, using (8), we can show that

$$
\begin{aligned}
E[m_t(b_L \mid L)] &= E\left(vec\left[\begin{pmatrix} 1 \\ z_{\bullet t} \end{pmatrix} (H_L' g_{\bullet t} - a_{P-L})' \right] \right) \\
&= vec \begin{pmatrix} (H_L' \alpha^g - a_{P-L})' \\ \alpha^z (H_L' \alpha^g - a_{P-L})' + \mathrm{B}^z \Omega_f \mathrm{B}^{g\prime} H_L \end{pmatrix} \neq 0
\end{aligned}, \tag{A.1}
$$

for any H_L and a_{P-L}. Thus, (i) holds. When $L = L^*$, there exists a unique $H_L = [I_{P-L^*}, -\Xi_2^{*\prime}]' \equiv \Xi^*$ such that $\mathrm{B}^z \Omega_f \mathrm{B}^{g\prime} \Xi^* = 0$. Let $\theta^* = vec[(\alpha_\Xi^*, \Xi_2^{*\prime})']$, where $\alpha_\Xi^* = \Xi^{*\prime} \alpha^g$. Then, from (A.1), we can see that θ^* is the unique solution of $E[m_t(b_L \mid L^*)] = 0$. This proves (ii). A simple example can provide an intuition for the result (iii). Consider a simple case with $P = Q = 3$, $L^* = 1$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$, $g_{\bullet t} = (g_{1t}, g_{2t}, g_{3t})'$, and

$$
\mathrm{B}^z \Omega_f \mathrm{B}^{g\prime} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix}; \ \Xi^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\gamma_1/\gamma_3 & -\gamma_2/\gamma_3 \end{pmatrix},
$$

assuming $\gamma_3 \neq 0$. Suppose now that for estimation, we choose $L = 2$ and $H_L' = (1, -\phi_1, -\phi_2)'$, where ϕ_1 is a unrestricted parameter, and $\phi_2 = -(\gamma_1 - \phi_1 \gamma_2)/\gamma_3$. Then, $\mathrm{B}^z \Omega_f \mathrm{B}^{g\prime} H_L$ = 0. Thus,

$E[vec\{(1,z'_{\bullet t})'(H'_L g_{\bullet t} - a_{P-L})'\}] = 0$, where $a_{P-L} = \alpha_1 - \phi_1\alpha_2 - \phi_2\alpha_3$. Since ϕ_1 is an unrestricted parameter, there are infinitely many matrices H_L (other than Ξ^*) that satisfy the moment conditions $E[m_t(b_L \mid L)] = 0$ when $L > L^*$. This result can be easily generalized.

We first prove Proposition 3, and later Proposition 1. We do so because Proposition 1 is in fact a corollary of Proposition 3.

Proof of Proposition 3: Consider the case with $L < L^*$. Then, by Lemma 1(i), there is no b_L that satisfies $E[m_t(b_L \mid L)] = 0$. This means that $\hat{b}_L \to_p b$ for some $b \neq \theta^*$, and, for some $d \neq 0$,

$$d_T(\hat{b}_L \mid L) = d_T(b^* \mid L) + o_p(1) \to_p d.$$

Thus, $c_T(\hat{b}_L \mid W_T(L), L)/T \to_p d'[W(L)]^{-1} d > 0$, and therefore, $c_T(\hat{b}_L \mid W_T(L), L) \to_p \infty$.

Consider now the case with $L = L^*$. By Lemma 1(ii), there is a unique solution θ^* for $E[m_t(b_L \mid L_o)] = 0$. Therefore, $\hat{b}_L \to_p \theta^*$ by Hansen (1982). Let $D_T(b_L \mid L) = \partial d_T(b_L \mid L)/\partial b'_L$. Then, using (7), we can show that, under Assumptions A-C,

$$D_T(\hat{b}_L \mid L^*) = \frac{1}{T}\Sigma_{t=1}^T \left(I_{P-L} \otimes \begin{pmatrix} 1 & (x_{\bullet t}^{L^*})' \\ z_{\bullet t} & z_{\bullet t}(x_{\bullet t}^{L^*})' \end{pmatrix} \right)$$

$$\to_p I_{P-L} \otimes \begin{pmatrix} 1 & (\alpha_{L^*}^x) \\ \alpha^z & \alpha^z(\alpha_{L^*}^x)' + \mathrm{B}^z\Omega_f(\mathrm{B}_{L^*}^x)' \end{pmatrix} \equiv D,$$

where D is a full-column matrix under Assumption D*. Define

$$A = \tilde{W}^{1/2}W^{-1/2}\left[I_{(P-L_o)(1+Q)} - (W')^{-1/2}D(D'W^{-1}D)^{-1}D'W^{-1/2}\right](W')^{-1/2}(\tilde{W}')^{1/2},$$

where $\tilde{W} = \tilde{W}(L^*)$, $W = W(L^*)$, and $\tilde{W}^{1/2}$ and $W^{1/2}$ are the triangular matrices from the Cholesky decomposition of S and G. Then, following Theorem 3 of Jagannathan and Wang (1996), we can show that, as $N \to \infty$,

$$c_T(\hat{b}_L \mid W_T(L^*), L^*) \to_d \Sigma_{j=1}^G \lambda_j \xi_j, \tag{A.2}$$

where $G = (P - L^*)(Q - L^*)$, $\lambda_1, \ldots, \lambda_G$ are the *positive* eigenvalues of A, and the ξ_j are independent $\chi^2(1)$ random variables. Observe that G is the degrees of overidentifying restrictions: the number of moment restrictions in $m_t(b_L \mid L)$ minus the number of parameters in b_L. That is,

$$G = (P - L^*)(1+Q) - (P - L^*)(1+L^*) = (P - L^*)(Q - L^*).$$

The minimizer of $c_T(b_L \mid \tilde{W}_T(L^*), L^*)$, $\tilde{b}_L^*$, is the optimal GMM estimator among the estimators based on the moment condition $E[m_t(b_L \mid L^*)] = 0$. When $\tilde{W}_T(L^*)$ is used for GMM, the matrix A reduces to:

$$A = I_{(P-L_o)(1+Q)} - (W')^{-1/2} D(D'W^{-1}D)^{-1} D'W^{-1/2},$$

because $\tilde{W} = W$. Observe that A is symmetric and idempotent. Thus, the eigenvalues of A are all ones or zeros. Since the rank of a matrix equals the number of its non-zero eigenvalues, and since $rank(A) = trace(A) = G$, we must have $\lambda_j = 1$, for all $j = 1,...,G$. Thus, by (A.2),

$$c_T(\tilde{b}_L^* \mid \tilde{W}_T(L^*), L^*) \to_d \Sigma_{j=1}^G \xi_j = \chi^2(G).$$

Proof of Proposition 1: Under Assumption D, $rank[\mathrm{B}^z \Omega_f \mathrm{B}^{g\prime}] = L^* = L_o$. Thus, the results follow from Proposition 3.

Proof of Proposition 2: We can complete the proof by showing that $\Pr(\hat{L} > L_o) \to 0$ and $\Pr(\hat{L} < L_o) \to 0$, as $T \to \infty$. We first show $\lim_{N\to\infty} \Pr(\hat{L} > L_o) = 0$. Observe that for $\hat{L}$ to be greater than L_o, $MS_T(L_o) - MS_T(L_a) > 0$ for some $L_a > L_o$. This implies that $\Pr(\hat{L} > L_o) \le \Pr[MS_T(L_o) - MS_T(L_a) > 0]$. Note also that

$$\begin{aligned}\Pr[MS_T(L_o) - MS_T(L_a) > 0] &= \Pr[c_T(\hat{b}_L^o \mid L_o) - c_T(\hat{b}_L \mid L_a) + f(N)(g(L_a) - g(L_o)) > 0] \\ &\le \Pr[c_T(\hat{b}_L^o \mid L_o) + f(T)(g(L_a) - g(L_o)) > 0] \to 0,\end{aligned}$$

because $(g(L_a) - g(L_o))$ is a fixed negative number, $f(T) = \ln(T) \to \infty$ as $T \to \infty$, and $c_T(\hat{b}_L^o \mid L_o)$ is a weighted χ^2 random variable that is bounded almost surely. Thus, $\Pr(\hat{L} > L_o) \to 0$.

We now show that $\lim_{T\to\infty} \Pr(\hat{L} < L_o) = 0$. For $\hat{L}$ to be smaller than L_o, there should exist some $L_a < L_o$ such that $MS_T(L_o) - MS_T(L_a) < 0$. But we can show

$$\Pr[MS_T(L_o) - MS_T(L_a) > 0] = \Pr\left[\frac{1}{T}c_T(\hat{b}_L^o \mid L_o) - \frac{1}{T}c_T(\hat{b}_L \mid L_a) + \frac{f(T)}{T}(g(L_a) - g(L_o)) > 0\right]$$
$$\to 0,$$

as $T \to \infty$. This is true for three reasons. First, $f(T)/T = \ln(T)/T \to 0$ as $T \to \infty$. Thus, the third term converges to zero. Second, $c_T(\hat{b}_L^o \mid L_o)/T \to_p 0$, because $c_T(\hat{b}_L^o \mid L_o)$ converges to a bounded random variable. Third, for $L_a < L_o$, there is no b_L such that $E[m_t(b_L \mid L_a)] = 0$. Let b_a be the minimizer of $p\lim_{T\to\infty} c_T(b_L \mid W_T(L_a), L_a)/T$. Since $E(m_t(b_a \mid L_a)) \neq 0$,

$$p\lim_{T\to\infty} d_T(\hat{b}_L \mid L_a) = p\lim_{T\to\infty} d_T(b_a \mid L_a) \neq 0.$$

Thus, $c_T(\hat{b}_L \mid L_a)/T$ converges in probability to a positive number. Accordingly, $\Pr(\hat{L} < L_o) \to 0$.

REFERENCES

Ahn, Seung C., Young H. Lee, and Peter Schmidt, 2007a, Stochastic frontier models with multiple time-varying individual effects, *Journal of Productivity Analysis* 27 (1), 1-12.

Ahn, Seung C., Young H. Lee, and Peter Schmidt, 2007b, Panel data models with multiple time-varying individual effects, mimeo, Arizona State University.

Ahn, Seung C., Stephan Dieckmann and Marcos F. Perez, 2007, Estimating the Common Factor in Credit Spreads, mimeo, Arizona State University.

Andersen, Torben G., and Bent E. Sørensen, 1996, GMM estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Business & Economic Statistics* 14, 328-352

Andrews, Donald W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.

Atkinson, Anthony .C., 1981 Likelihood ratios, posterior odds and information criteria, *Journal of Econometrics* 16, 15-20

Bai, Jusan and Serena Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191-221.

Bai, Jusan, 2003, Inferential theory for factor models of large dimensions, *Econometrica* 71, 135-171.

Black, Fisher, Michael C. Jensen, and Myron S. Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets*, Praeger, New York.

Brown, Stephen J. and Mark I. Weinstein, 1983, A new approach to testing asset pricing models: The bilinear paradigm, *Journal of Finance* 38, 711-743.

Brown, Stephan J., 1989, The number of factors in security returns, *Journal of Finance* 44(5), 1247-1262.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Chamberlain, Gary and Michael Rothschild, 1983, Arbitrage, factor Structure, and mean variance analysis on large asset markets, *Econometrica* 51, 1281-1304.

Chen, Nai Fu, Richard Roll and Stephen Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 368-403.

Connor, Gregory, 1984, A unified beta pricing theory, *Journal of Economic Theory* 34, 13-31.

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

Connor, Gregory and Robert A. Korajczyk, 1993, A test for the number of factors in an approximate factor model. *Journal of Finance* 48, 1263-1291.

Cragg John and Stephen Donald, 1996, On the asymptotic properties of LDU based test of the rank of a matrix, *Journal of the American Statistical Association* 91, 1301-1309.

Cragg John and Stephen Donald, 1997, Inferring the rank of a matrix, *Journal of Econometrics* 76, 223-250

Donald Stephen G., Fortuna Natercia, and Vladas Pipiras, 2005, On the rank estimation in symmetric matrices: the case of indefinite estimators, CEMPRE Working Paper

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: empirical tests, *Journal of Political Economy* 71, 607-636.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Ferson, Wayne, and Stephen R. Foerster, 1994, Finite sample properties of the generalized methods of moments tests of conditional asset pricing models, *Journal of Financial Economics* 36, 29-56.

Gill, Len, and Arthur Lewbel, 1992, Testing the rank and definiteness of estimated matrices with application to factor, state-space and ARMA models, *Journal of the American Statistical Association*, 87, 766-776.

Grinblatt, Mark and Sheridan Titman, 1985, Approximate factor structures: Interpretations and implications for empirical tests, *Journal of Finance* 40, 1367-1373.

Gorman, William M., 1981, Some Engel curves, in *Essay in the Theory and Measurement of Consumer Behavior in Honor of Sir Richard Stone*, ed. by A. Deaton, New York: Cambridge University Press.

Gregory, Allan W., and Allen C. Head, 1999, Common and country-specific fluctuations in productivity, investment, and the current account, *Journal of Monetary Economics* 44, 423.

Hannan, Edward J., 1980, The estimation of the order of an ARMA process, *Annals of Statistics* 8, 1071-1081.

Hannan, Edward. J., 1981, Estimating the dimension of linear system, *Journal of Multivariate analysis* 11, 459-473.

Hannan, Edward J., and B. G. Quinn, 1979, The determinants of the order of an autoregression, *Journal of the Royal Statistical Society* 41, 190-195.

Hansen, Lars P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1054.

Hinkley, David V., 1977, Jackknifing in unbalance situations, *Technometrics* 19, 285-292.

Jones, Christopher S., 2001, Extracting factors from heteroskedastic returns, *Journal of Financial Economics* 62, 293-325.

Jöreskog, Karl G., 1967, Some contributions to maximum likelihood factor analysis, *Psychometrika* 34, 183-202.

Kan, Raymond, and Chu Zhang, 1999, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 204-235.

Lehmann, Bruce N., and David M. Modest, 1988, The empirical foundations of the arbitrage pricing theory, *Journal of Financial Economics* 21, 213-254.

Lewbel, Arthur, 1991, The rank of demand system: Theory and nonparametric estimation, *Econometrica* 59, 711-730.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55 (3), 703-708

Newey, Whitney K., and Kenneth D. West, 1994. Automatic Lag Selection in Covariance Matrix Estimation *The Review of Economic Studies*, 61, 631-653

Nishii, Ryuie, 1988, Maximum likelihood principle and model selection when the true model is unspecified, *Journal of Multivariate Analysis*, 27, 392-403

Roll, Richard W. and Stephan A. Ross, 1980, An empirical investigation of the arbitrage pricing theory, *Journal of Finance* 35, 1073-1103.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.

Stock, James H., and Mark Watson, 2005, Implication of dynamic factor models for VAR analysis, NBER working paper 11467.

White, Halbert L., 1980, A heteroskedasticity-consistent covariance matrix estimator and direct test for heteroskedasticity, *Econometrica* 48, 817-838

White, Halbert L., 1999, *Asymptotic Theory for Econometricians*: Academic Press, San Diego, California.

TABLE 1
Size of the J Test

1,000 random samples are generated by a three-factor model. For each sample, the J statistic is computed and the null hypothesis of $L_0 = 3$ is tested. The reported are the percentages of rejection at 5% of significance level. The Newey-West estimator is used to compute the weighting matrix for the cases of autocorrelation. The abbreviation BW is the value of bandwidth used.

NO AUTOCORRELATION					
T	N	BW	Number of Instruments (Q)		
			4	5	6
500	12	0	4.30%	5.10%	7.10%
1000	12	0	4.70%	4.90%	5.60%
T	N	BW	Number of Instruments (Q)		
			6	8	9
500	15	0	7.20%	5.30%	5.20%
1000	15	0	6.40%	4.90%	4.85%
T	N	BW	Number of Instruments (Q)		
			8	9	12
500	25	0	6.80%	4.90%	33.2%
1000	25	0	5.90%	5.00%	32.50%
AUTOCORRELATION					
T	N	BW	Number of Instruments (Q)		
			4	5	6
500	12	3	4.00%	4.30%	5.70%
1000	12	3	4.50%	5.00%	4.60%
500	12	0	4.50%	6.10%	7.60%
1000	12	0	4.70%	6.00%	5.80%
T	N	BW	Number of Instruments (Q)		
			6	8	9
500	15	3	3.20%	2.50%	2.90%
1000	15	3	5.00%	2.60%	2.90%
500	15	0	7.20%	5.80%	5.10%
1000	15	0	6.20%	5.00%	4.60%
T	N	BW	Number of Instruments (Q)		
			8	9	12
500	25	3	0.00%	0.00%	0.00%
1000	25	3	0.20%	0.00%	0.74%
500	25	0	7.40%	5.20%	57.00%
1000	25	0	5.90%	5.30%	47.20%

TABLE 2

Estimation by the Sequential Hypothesis Testing Method

The sequential hypothesis testing method is used to estimate the number of factors in the data generated with three factors (L_o = 3). Five, eight and nine instrumental variables are used for the data with N = 12, 15, and 25, respectively. For the cases of autocorrelation, we use the Newey-West estimator. The total number of simulations is 1,000.

NO AUTO CORRELATION

T	N	BANDWIDTH	NUMBER OF FACTORS					AVERAGE
			≤ 2	3	4	5	6	
500	12	0	0%	94.10%	5.90%	0.00%	0.00%	3.059
500	15	0	0%	94.70%	3.70%	1.10%	0.50%	3.074
500	25	0	0%	93.20%	4.40%	1.50%	0.08%	3.052
1000	12	0	0%	94.20%	5.80%	0.00%	0.00%	3.058
1000	15	0	0%	95.10%	3.40%	1.30%	0.02%	3.055
1000	25	0	0%	94.40%	3.40%	1.40%	0.20%	3.050

AUTOCORRELATION

T	N	BANDWIDTH	NUMBER OF FACTORS					AVERAGE
			≤ 2	3	4	5	6	
500	12	0	0%	94.60%	5.40%	0.00%	0.00%	3.054
500	15	0	0%	94.20%	4.30%	0.90%	0.60%	3.079
500	25	0	0%	94.10%	3.00%	1.30%	0.60%	3.044
1000	12	0	0%	94.70%	5.30%	0%	0%	3.053
1000	15	0	0%	95.40%	3.70%	0.90%	0.00%	3.055
1000	25	0	0%	94.60%	3.40%	1.20%	0.40%	3.058
500	12	3	0%	95.70%	4.30%	0.00%	0.00%	3.043
500	15	3	0%	97.50%	1.60%	0.80%	0.10%	3.035
500	25	3	100%	0.00%	0.00%	0.00%	0.00%	1.000
1000	12	3	0%	95.00%	5.00%	0%	0%	3.050
1000	15	3	0%	97.40%	1.40%	0.80%	0.40%	3.042
1000	25	3	0%	92.30%	5.70%	3.00%	0.00%	3.147

TABLE 3

Estimating the number of Factors by the Model Selection Criterion Method

The model selection criterion method is used to estimate the number of factors for the data generated with three factors (L_0 = 3). Five, eight and nine instrumental variables are used for the data with N = 12, 15, and 25, respectively. For the cases of autocorrelation, we use the Newey-West estimator. The total number of simulations is 1,000. The Schwartz Information Criterion (BIC) is used.

NO AUTO CORRELATION

T	N	BANDWIDTH	NUMBER OF FACTORS					AVERAGE
			≤ 2	3	4	5	6	
500	12	0	0.00%	97.00%	3.00%	0.00%	0.00%	3.03
500	15	0	0.00%	99.10%	0.90%	0.00%	0.00%	3.01
500	25	0	0.00%	99.90%	0.10%	0.00%	0.00%	3.00
1000	12	0	0.00%	97.80%	2.20%	0.00%	0.00%	3.02
1000	15	0	0.00%	99.70%	0.03%	0.00%	0.00%	2.99
1000	25	0	0.00%	99.80%	0.20%	0.00%	0.00%	3.00

AUTOCORRELATION

T	N	BANDWIDTH	NUMBER OF FACTORS					AVERAGE
			≤ 2	3	4	5	6	
500	12	0	0.00%	96.90%	3.10%	0.00%	0.00%	3.03
500	15	0	0.00%	98.90%	1.10%	0.00%	0.00%	3.01
500	25	0	0.00%	99.70%	0.10%	0.00%	0.00%	3.00
1000	12	0	0.00%	97.70%	2.30%	0.00%	0.00%	3.02
1000	15	0	0.00%	99.60%	0.40%	0.00%	0.00%	3.00
1000	25	0	0.00%	99.90%	0.10%	0.00%	0.00%	3.00
500	12	3	0.00%	98.00%	2.00%	0.00%	0.00%	3.02
500	15	3	0.00%	99.80%	0.20%	0.00%	0.00%	3.00
500	25	3	100.00%	0.00%	0.00%	0.00%	0.00%	1.00
1000	12	3	0.00%	98.10%	1.90%	0.00%	0.00%	3.02
1000	15	3	0.00%	99.90%	0.10%	0.00%	0.00%	3.00
1000	25	3	9.20%	90.80%	0.00%	0.00%	0.00%	2.82

TABLE 4

Effects of Changing Partitions

PANEL A

A <u>*single*</u> data set with T = 1,000 is generated from a three-factor model. The number of factors is estimated by sequential testing and model selection criterion methods. This estimation is conducted for 100 randomly chosen partitions of the response variables. For the cases of autocorrelation, the Newey-West estimator with bandwidth = 3 is used for the sequential hypothesis testing method. Zero bandwidth is used for the model selection criterion method. The term "yes" in the "auto" column indicates a case of autocorrelation, and "no" indicates the case of no autocorrelation.

SEQUENTIAL HYPOTHESIS TESTING

N	auto	NUMBER OF FACTORS				AVERAGE
		≤ 2	3	4	5	
12	no	0.00%	97.00%	3.00%	0.00%	3.03
15	no	0.00%	89.00%	10.00%	0.00%	3.07
25	no	0.00%	66.60%	33.30%	0.00%	3.33
12	yes	0.00%	97.00%	3.00%	0.00%	3.03
15	yes	0.00%	91.00%	8.00%	1.00%	3.10
25	yes	0.00%	65.00%	32.00%	3.00%	3.38

MODEL SELECTION CRITERION

N	auto	NUMBER OF FACTORS				AVERAGE
		≤ 2	3	4	5	
12	no	0.00%	97.00%	3.00%	0.00%	3.03
15	no	0.00%	91.00%	8.00%	1.00%	3.10
25	no	0.00%	83.00%	15.00%	2.00%	3.19
12	yes	0.00%	97.00%	3.00%	0.00%	3.03
15	yes	0.00%	91.00%	8.00%	1.00%	3.10
25	yes	0.00%	82.00%	16.00%	2.00%	3.20

PANEL B

1,000 random samples are generated from a three-factor model. For each generated sample, the number of factors is estimated by applying the model selection criterion method to 100 randomly chosen partitions of response variables. The estimated number of factors for each sample is the number estimated the most frequently.

MODEL SELECTION CRITERION

N	auto	NUMBER OF FACTORS				AVERAGE
		≤ 2	3	4	5	
12	no	0.00%	100.00%	0.00%	0.00%	3.00
15	no	0.00%	100.00%	0.00%	0.00%	3.00
25	no	0.00%	100.00%	0.00%	0.00%	3.00
12	yes	0.00%	100.00%	0.00%	0.00%	3.00
15	yes	0.00%	100.00%	0.00%	0.00%	3.00
25	yes	0.00%	100.00%	0.00%	0.00%	3.00

TABLE 5

Effects of Weak Factor in One Random Sample

A single data set with T = 1,000 is generated from a three-factor model using different signal to noise ratios for the second common component. The variances of the response variables explained by three factors are held constant. The number of factors is estimated by the model selection criterion method. The method is applied to 100 randomly chosen partitions of response variables.

PANEL A: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 1.00

N	NUMBER OF FACTORS					AVERAGE
	≤ 1	2	3	4	5	
12	0.00%	4.00%	91.00%	5.00%	0.00%	3.01
15	0.00%	0.00%	90.00%	9.00%	1.00%	3.11
25	0.00%	0.00%	75.00%	24.00%	1.00%	3.26

PANEL B: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 0.50

N	NUMBER OF FACTORS					AVERAGE
	≤ 1	2	3	4	5	
12	0.00%	24.00%	71.00%	5.00%	0.00%	2.81
15	0.00%	6.00%	82.00%	12.00%	1.00%	3.11
25	0.00%	14.00%	66.00%	19.00%	1.00%	3.07

PANEL C: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 0.35

N	NUMBER OF FACTORS					AVERAGE
	≤ 1	2	3	4	5	
12	0.00%	35.00%	60.00%	4.00%	0.00%	2.66
15	0.00%	23.00%	67.00%	10.00%	0.00%	2.87
25	0.00%	13.00%	59.00%	24.00%	4.00%	3.19

PANEL C: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 0.25

N	NUMBER OF FACTORS					AVERAGE
	≤ 1	2	3	4	5	
12	0.00%	75.00%	25.00%	0.00%	0.00%	2.22
15	0.00%	60.00%	39.00%	10.00%	0.00%	2.77
25	0.00%	68.00%	30.00%	2.00%	0.00%	2.34

TABLE 6

Effects of Weak Factors in 1,000 Random Samples

1,000 random samples are generated from a three-factor model with different signal to noise ratios of the second common component. Each sample contains 1,000 time series observations. For each data set, the variances of the response variables explained by three factors are held constant. The number of factors is estimated by applying the model selection criterion to 100 randomly chosen partitions of response variables. The estimated number of factors for each sample is the number estimated the most frequently.

PANEL A: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 1.00

N	METHOD	NUMBER OF FACTORS				AVERAGE
		≤ 1	2	3	4	
12	MSC	0.00%	0.00%	100.00%	0.00%	3.00
15	MSC	0.00%	0.00%	100.00%	0.00%	3.00
25	MSC	0.00%	0.00%	100.00%	0.00%	3.00

PANEL B: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 0.50

N	METHOD	NUMBER OF FACTORS				AVERAGE
		≤ 1	2	3	4	
12	MSC	0.00%	0.90%	99.10%	0.00%	2.99
15	MSC	0.00%	0.00%	100.00%	0.00%	3.00
25	MSC	0.00%	0.00%	100.00%	0.00%	3.00

PANEL C: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 0.35

N	METHOD	NUMBER OF FACTORS				AVERAGE
		< 1	2	3	4	
12	MSC	0.00%	38.30%	61.70%	0.00%	2.62
15	MSC	0.00%	6.50%	93.50%	0.00%	2.94
25	MSC	0.00%	0.00%	99.80%	0.20%	3.00

PANEL D: SIGNAL TO NOISE FOR COMMON COMPONENT 2 = 0.25

N	METHOD	NUMBER OF FACTORS				AVERAGE
		< 1	2	3	4	
12	MSC	0.00%	89.10%	10.90%	0.00%	2.11
15	MSC	0.00%	67.80%	32.30%	0.00%	2.33
25	MSC	0.00%	46.50%	53.50%	0.00%	2.54

TABLE 7
Empirical Application

We apply the model selection criterion method to three different groups of portfolios constructed from US stock market from 1929 to 2006. The 15 and 25 portfolios in each group are constructed randomly while each portfolio has the same number of assets. The 10 momentum portfolio was obtained from K. French data library and it has roughly the same number of assets per portfolio. PANEL A, shows results of the estimation of the number of factors for 100 randomly chosen partitions. The heteroskedasticity-robust variance matrix is used as weighting matrix. PANEL B includes results for testing how many factors are captured by Chen-Roll-Ross 1986 (CRR) factors and Fama and French (FF) 1996 factors, in a sub sample from 1978 of 342 observations..

PANEL A: Full sample

Portfolio Used	NUMBER OF FACTORS					Average
	1	2	3	4	5	
10 Momentum Portfolios	0.00%	8.00%	71.00%	21.00%	-	3.13
15 Random Portfolios	0.00%	35.00%	50.00%	6.00%	1.00%	2.44
25 Random Portfolios	8.00%	63.00%	25.00%	4.00%	1.00%	2.26
10 from the 25 Portfolios	16.00%	49.00%	31.00%	4.00%	-	2.23
10 from the 25 Portfolios	37.00%	40.00%	19.00%	4.00%	-	1.90

PANEL B: Number of Factors with Instruments (sub-sample from 1978)

Portfolio Used	NUMBER OF FACTORS					Average
	1	2	3	4	5	
10 Momentum Portfolios	0.00%	8.00%	78.00%	14.00%	-	3.06
CRR no Inflation Factors	100%	0%	-	-	-	1.00
CRR All Factors	100%	0%	0%	0%	-	1.00
FF Factors	5%	95%	-	-	-	1.95
CRR and FF Factors	29%	71%	0%	0%	0%	1.71