GMM Estimation of the Number of Latent Factors

Seung C. Ahn, Marcos Fabricio Perez[a]

September 17, 2008

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 find that the international stock returns are explained by one strong global factor. This factor seems to be highly correlated with the US stock market factors. This result can be interpreted as evidence for market integration. We also find two weak factors related mostly with the European and the Americas markets.

---

[a] Seung C. Ahn, is affiliated to the Department of Economics of W.P. Carey School of Business at Arizona State University and Marcos Fabricio Perez to the Finance Area of the School of Business and Economics at Wilfrid Laurier University. We would like to thank Stephan Dieckmann, Federico Nardari, Alex Horenstein, Michael Melvin and Andriy Shkilko for very useful discussions about this paper, as well as seminar participants at Arizona State University, Wilfrid Laurier University, the Northern Finance Association and the Canadian Econometric Study Group for helpful comments.

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 find that the international stock returns are explained by one strong global factor. This factor seems to be highly correlated with the US stock market factors. This result can be interpreted as evidence for market integration. We also find two weak factors related mostly with the European and the Americas markets.

Keywords: factor models, GMM, number of factors, international stock market

1. INTRODUCTION

Many economic and financial theories are based on linear factor models. The most used 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).

Estimation of the true number of factors is crucial for the theoretical and empirical validity of a model. For example in the case of the APT model, if too few factors are used in the estimation, the estimated factor prices will be inconsistent. In contrast, if too many factors are used, that is, if some of the used factors are not correlated with the returns, the beta risk of the useless factor can look priced even if the sample size is large (Kan and Zhang, 1999). Also. in factor models, it is important to determine what observable macroeconomic and/or financial variables are related to the unobservable factors, since this will 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[2] (*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.[3] 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 20. 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 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, 2008). Third, the method is computationally easy to implement. All necessary procedures are based on closed-

---

[2] 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.

[3] 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).

form solutions, and thus, we do not require non-linear optimization. Any software that can estimate multiple equations models can be used[4]. Fourth, the method allows for cross-section and time series heteroskedasticity and time series autocorrelation of the idiosyncratic components and it 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 international stock markets comovements. Our empirical results conclude that one strong global factor explains the comovement of international stock markets. Interestingly, this factor seems to be correlated with US stock market factors. This can interpreted as evidence for some degree of market integration. We also find evidence of two weak factors mostly related with the European and Americas markets

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 international 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

---

[4] A step by step practitioners guide for the method can be found at http://www.public.asu.edu/~mfperez. Also Gauss codes are available to download.

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 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 $\mathrm{B} = (\beta_1, \beta_2, ..., \beta_N)'$. We can assume that $E(f_t) = 0$ without loss of generality since we include a non-zero vector of intercepts into the model.

Since our method to estimate the number of factors (*L*) is an application of GMM, 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) $T^{-1/2}\Sigma_{t=1}^{T}w_t \rightarrow_d N\left(0_{(N+L_o)\times 1}, \Lambda\right)$, as $T \rightarrow \infty$, 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 $\Lambda = p\lim_{T\rightarrow\infty} T^{-1}\Sigma_{s=1}^{T}\Sigma_{t=1}^{T}E(w_t w_s')$.

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 and it can be relaxed without altering our results. The required assumption is that $T^{-1}\Sigma_{t=1}^{T}f_t f_t' \rightarrow_p \Omega_f$ as $T \rightarrow \infty$. 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[5]. Thus, with (C2), the model (2) is an exact factor 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. Also, even if the errors are cross sectionally correlated, the method can be used to estimate the number of factors by grouping the response variables appropriately (e.g., portfolios), as explained in section 4.1

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.

---

[5] 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.

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.

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

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

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

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 since 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 the computational difficulty of 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.[6]

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

$$\underset{(P\times 1)}{g_{\bullet t}}=\begin{pmatrix}\underset{((P-L)\times 1)}{y_{\bullet t}^{P-L}}\\ \underset{(L\times 1)}{x_{\bullet t}^{L}}\end{pmatrix}=\begin{pmatrix}\underset{((P-L)\times 1)}{\alpha_{P-L}^y}+\underset{((P-L)\times L_o)}{\mathrm{B}_{P-L}^y}\underset{(L_o\times 1)}{f_t}+\underset{((P-L)\times 1)}{\varepsilon_{P-L,\bullet t}^y}\\ \underset{(L\times 1)}{\alpha_L^x}+\underset{(L\times L_o)}{\mathrm{B}_L^x}\underset{(L_o\times 1)}{f_t}+\underset{(L\times 1)}{\varepsilon_{L,\bullet t}^x}\end{pmatrix}, \tag{6}$$

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

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

---

[6] 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.

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 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:

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

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:

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

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 $\lim_{T\to\infty} Var\left(\sqrt{T} d_T(b_L \mid L_o)\right)$. 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 $c_T(b_L \,|\, \tilde{W}_T(L_o), L_o)$. 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 $W_T(L)$, $c_T(\hat{b}_L \,|\, W_T(L), L) \to_p \infty$ for any $L < L_o$ and $c_T(\hat{b}_L^o \,|\, W_T(L_o), L_o) \to_d \Upsilon$, where Υ is a weighted average of independent $\chi^2(1)$ random variables. In addition, $c_T(\tilde{b}_L^o \,|\, \tilde{W}_T(L_o), L_o) \to_d \chi^2[(P - L_o)(Q - L_o)]$.

The proof of Proposition 1 is given in the appendix. The distribution of $c_T(\hat{b}_L \,|\, W_T(L), L)$ 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 $c_T(\tilde{b}_L^o \,|\, \tilde{W}_T(L_o), L_o)$ statistic, which is simply the over identifying 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 over identifying 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 over identifying 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 over identifying 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, $\hat{L} \rightarrow_p L_o$.

The proof of Proposition 2 is given in the appendix[7]. 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

[7] The proof is an 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.

practice, the selection criterion method appears to be an attractive alternative to the sequential method. In our Monte Carlo simulations (section 5) we compare the performance of the BIC criterion with the following criterions :

Akaike Information (AIC): $f(T)=1$, and $g(T)=2(P-L)(Q-L)$

Schwarz Criterion 2 (BIC2): $f(T)=\log(T)$, and $g(T)=(P-L)(Q-L)$.

Schwarz Criterion 3 (BIC3): $f(T)=\ln(T)$, and $g(T)=(P-L)(Q+1)$.

It can be shown that Proposition 2 holds using the modifications to the Schwarz Criterion labeled BIC2 and BIC3. AIC criterion is commonly used, but it leads to inconsistent estimation of the number of factors L_o

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

In the appendix (Lemma A.1) we shown that when $L=L^*$, a unique vector θ^* exists such that $E[m_t(\theta^*\,|\,L^*)]=0$. Let $\tilde{W}_T(L^*)$ be a consistent estimator of $\lim_{T\to\infty}Var\left(\sqrt{T}d_T(\theta^*\,|\,L^*)\right)$; and let $\hat{b}_L^*$ and $\tilde{b}_L^*$ be the minimizers of $c_T(b_L\,|\,W_T(L^*),L^*)$ and $c_T(b_L\,|\,\tilde{W}_T(L^*),L^*)$, 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 $W_T(L)$, $c_T(\hat{b}_L\,|\,W_T(L),L)\to_p\infty$. In contrast, $c_T(\hat{b}_L^*\,|\,W_T(L^*),L^*)\to_d\Upsilon$, where Υ is a weighted average of independent $\chi^2(1)$ random variables. In addition, $c_T(\tilde{b}_L^*\,|\,\tilde{W}_T(L^*),L^*)\ \to_d\ \chi^2[(P-L^*)(Q-L^*)]$.

Since the partition of $g_{\bullet t}$ and $z_{\bullet t}$ is arbitrary, the rank of $B^z\Omega_f B^{g\prime}$ 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. The estimator we propose is the number which is most often estimated from the estimation with different partitions. From now on, we will refer to this estimator as "highest frequency estimator" (HFE).

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, $(NT)^{-1} \Sigma_{t=1}^{T} \Sigma \left| \tau_{ii',tt} \right| \leq \left| \tau_{ii'} \right|$ for some $\tau_{ii'}$ and for all t, and $N^{-1} \Sigma_{i=1}^{N} \Sigma_{i'=1}^{N} \left| \tau_{ii'} \right| \leq M$ for some positive number M, for all N.

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

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

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

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

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

where the symbols with overhead bar are defined similarly to $\bar{u}_{j,t}$. By (12-1)-(12-2) and the fact that the variables in $f_{j,t}^{loc}$ are group-specific, we can show that $\bar{\varepsilon}_{j,t}$ 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 $\mathrm{B}^{glo} = \lim_{N\to\infty}\left(\bar{\beta}_1^{glo}, \bar{\beta}_2^{glo}, ..., \bar{\beta}_n^{glo}\right)'$.

4.2. GMM Estimation with Observable Instruments

When some observable variables are potentially correlated with the latent factors, we could use them to estimate L_o, or to test how many of them are indeed correlated with the factors. This test will allow the researcher to give economics meaning to the latent factors and to test if the variables proposed by economic or financial models are in fact correlated with the latent 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

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

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. Simulation Design

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

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

where the f_{kt} (k = 1, 2, 3) are the common factors and $c_{k,it}$ are the common components of the model. Our benchmark model is the three-factor model of Fama and French (1993): EMR (excess market return), SMB, and HML.[8] 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:

$$var(\bar{\beta}_1 \times EMR) = 21.66; var(\bar{\beta}_2 \times SMB) = 4.51; var(\bar{\beta}_3 \times HML) = 1.31.$$

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} range between 1.21 and 3.79, with an average of 2.016. Thus, the variances of the first and second common components at the means of betas

[8] 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 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.

are more than twice as great as the average variance of the idiosyncratic component, while the variance of the third common component (1.31) 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 approximately 10.74 , 2.23, 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 $\text{var}(v_{it})$ such that $\text{var}(\varepsilon_{it}) = \sigma_i^{FF}$. All the estimated values of ρ_i from equation (15) are smaller than 0.1 evidencing very small serial correlation. Using such small values for ρ_i in our simulations give almost identical results that in the non-autocorrelation case. Instead, we generate data using ρ_i=0.5, a value that better help us evaluate the performance of the method under serial correlation over time. Errors generated in this way are cross-sectional heteroskedastic and serially correlated over time

We also include simulations where data is generated with just one factor: $r_{it} = \alpha_i + \beta_{i1} f_{1t} + \varepsilon_{it} = \alpha_i + c_{1,it} + \varepsilon_{it}$. In this case the benchmark factor used for data generation is HML factor. That is, the only common component in this case ($c_{1,it}$) will be generated matching the moments of the HML factor. We use the HML factor since it has the smallest signal to noise ratio.

As discussed in section 3, to obtain consistent estimates 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 $\alpha_{500} = 0.05$.[9] 1,000 different sets of randomly generated portfolio returns are used for simulations. The weighting matrix used in GMM is Newey-West (1987). Note that this weighting matrix collapses to the heteroskedasticity-robust weighting matrix of White (1980) when bandwidth is equal to zero. Finally, as described in section 3 our methodology requires the partition of the response variables in two groups (groups $g_{\bullet t}$ and $z_{\bullet t}$ in the notation of section 2). The *Q* portfolios in group $z_{\bullet t}$

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

will be our instruments in the estimation. All the reported simulation results in the following section use half of the response variables ($N/2$) as instruments[10]. Results using other number of instruments are available from the authors upon request[11].

5.2. Changing Partitions and Highest Frequency Estimator (HFE)

As described in sections 2 and 3 our estimation procedure requires the partition of the response variables in two groups. Since this partition is arbitrary, the estimation result could change depending on the choice of partition The purpose of this first simulation exercise is to evaluate the effect in the estimation of changing the portfolios included on each partition group. To accomplish this objective we generate one set of N portfolio returns with three factors following (15), and then we partition the N portfolios into two randomly selected groups. We estimate the number of factors using the sequential hypothesis testing (SHT) and the model selection criterion methods (MSC) for this partition. We repeat the estimation procedure for 100 randomly selected partitions of the generated N portfolios.

We report in Table 1 the number of times we estimate 1, 2, 3, 4 and 5 or more factors for the 100 partitions described before. Results are reported for different values of N and T. Panel A includes results using the sequential hypothesis testing methodology and Panel B using the model selection criterion with penalty function BIC2[12]. We also present the results graphically to facilitate interpretation.

We are aware that this is a one sample exercise, we include it here in order to motivate the use of the highest frequency estimator and not to evaluate the small sample performance of the method. However, a researcher using the our method in empirical work will face results as the ones presented in Table 1.

In Table 1, using SHT the true number of factors (three) is always estimated with the highest frequency, with frequencies ranging from 81% to 98%. Similar results are obtained using MSC except for the case of $N=25$ $T=250$ where the method fails to the correctly estimate the number

[10] Results are not significantly sensible to the number of instruments used as long as they are close to $N/2$

[11] Many studies find that the GMM estimators computed with too many instruments and small data are often biased (see, for example, Andersen and Sørensen, 1996) 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.

[12] Result using other penalty functions are omitted to save space and they are available from the authors. The main conclusion do not change by including these results.

of factors. This experiment confirms that the GMM estimation could change depending on the partition we use. The test is more sensitive to the partitions used in GMM when N is large and T is small.

Given these results we propose that the estimation should be repeated 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.[13] The estimator we propose is the "highest frequency estimator" (HFE), which is simply the number of factors most often estimated from 100 different partitions. The remaining part of the Monte Carlo simulation is devoted to investigate the finite sample properties of the HFE under different specification of the data generating process.

5.3 Finite Sample Properties of the Highest Frequency Estimator

In order to investigate the finite-sample properties of HFE, 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. Our "Highest Frequency Estimator", for each Monte Carlo sample, will be the number of factors most often estimated from the 100 different partitions.

The first objective of our Monte Carlo simulations is to evaluate the performance of HFE when the idiosyncratic errors are cross-sectionally heteroskedastic, but not autocorrelated. Results presented in Table 2 include the average number of factors estimated over the 1,000 simulations using sequential hypothesis testing (SHT), and using the model selection criterion method with penalty functions BIC, BIC2, BIC3 and AIC. We also report estimation results using Bai and Ng (2002) methodology with penalty functions IC1, IC2, IC3, PC1, PC2 and PC3. Bai and Ng 2002 methodology is designed for data sets with large number of time series and cross sectional observations (i.e. large N and T). We include these results in order to evaluate for what combinations of N and T the HFE will be preferred over the Bai and Ng method. Table 2, Panel A includes results for data generated using 3 factors and Panel B for data generated using one factor.

---

[13] We also performed the same experiment using all possible partitions of portfolio returns into two groups. Results do not differ significantly with the ones presented just using 100 random specifications of the groups.

When data is generated with 3 factors, HFE estimates in average the correct number of factors for almost all combinations of N and T, independently of the penalty function used. Sequential hypothesis testing method and model selection criterion with penalty BIC2 are the more accurate methods. As expected the accuracy of the HFE improves as the number of time series observations (T) increases. The HFE method just fails estimating the number of factors when $N=25$ and $T=250$ using penalty functions BIC2, BIC3 and AIC. As expected the method deteriorates as N increases for small values of T. Contrarily, Bai and Ng methods are very accurate when $N=25$. No surprisingly the precision of their method decays as N get smaller, failing to estimate the number of factors for values of N smaller than 15. Similar results can be obtained for the case data is generated with one factor as showed in Table 2 Panel B.

The objective of our second experiment is to evaluate the performance of the HFE under the presence of autocorrelated idiosyncratic errors. We generate data with idiosyncratic errors following an AR(1) process with ρ_i=0.5 as described in section 5.1. White covariance matrix (which is asymptotically not an optimal choice) is used for the case of model selection criterion. As explained in section 3, model selection criterion does not require the use the optimal GMM estimator. In unreported experiments we use Newey-West covariance matrix, and in general results deteriorate specially for the cases with N = 15 and 25. 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[14]. For the case of sequential hypothesis testing we require to use the optimal GMM estimator. We report results Newey-West matrix with bandwidth 3[15]. We also report results using the non-optimal weighting matrix (bandwidth=0) in order to check the distortions of the estimation using sequential hypothesis testing when an non optimal weighting matrix is used.

Results using data generated with 3 factors and autocorrelated idiosyncratic errors are reported in Table 3. As expected the sequential hypothesis testing method using non-optimal GMM over estimates the true number of factors. Results significantly improve using Newey West methods with bandwidth=3 for almost all cases. The only exemptions are $N=15$, $T=250$;

[14] We also investigated the performance of other non optimal weighting matrixes. For example the case of a block diagonal weighting matrices, or equivalently the estimation of the system equation by equation. The results are similar in magnitude and computationally faster specially when N=25.
[15] Several other values for bandwidth were tested in our simulations. In general results are similar for small bandwidth choices. Results deteriorate for bandwidth levels bigger than 3, specially for $N=15$ and 25.

N=25, T=250 and *N=25, T=500*. We conclude that in the case of the sequential hypothesis testing method large samples are required to analyze a large number of portfolios ($N \geq 15$) when idiosyncratic errors are autocorrelated. For the case of model selection criterion, results are quite robust to time series autocorrelation. We estimate the correct number of factors for almost all combinations of *N* and *T*. BIC3 seems to be the most accurate penalty function for autocorrelated errors, but the differences with the other BIC penalty functions is very small. The only exemption to the good performance of the model selection criterion is the case *N=25, T=250* where HFE underestimates the number of factors. The inconsistent penalty function AIC in general over estimates the number of factors for all combinations of *T* and *N*.. Similarly that in the *iid* case, Bai and Ng methods are very accurate for *N=25*, but for *N≤15* the HFE is more precise.

The last part of our simulation exercise tries to evaluate the performance of the HFE when a factor explains a very small proportion of the total variation of the response variable. We will call such factor a *weak* factor[16]. 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 10.74 , 2.23, and 0.65, respectively. To generate the data with one weak factor, 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 idiosyncratic errors remain constant. In this experiment, we report results using data generated with a SNR of : 0.30 and 0.20 for the SMB factor[17].

In order to evaluate the effect of changing partitions in the presence of a weak factor first we proceed similarly that in the one sample experiment reported in Table 1. We generate one set of *N* portfolio returns with three factors following (15), and then we partition the *N* portfolios into

[16] A weak factor can also be interpreted as a factor that influences only a subset of the response variables. This is related with Assumption *D* explained in section 2. Theoretically when assumption D does not hold, our methods tend to underestimate the number of factors as stated in Proposition 3.
[17] We also performed simulations with higher SNR. Results are very similar to the case data is generated with 3 strong factors presented in Table 2.

two randomly selected groups. We estimate the number of factors for this partition. We repeat the estimation procedure for 100 randomly selected partitions. Results are presented in Table 4[18].

For the case SNR=0.30 (Panel A) the model selection criterion method still estimates the correct number of factors (3 factors) with the highest frequency. However if we compare this results with the ones presented in Table 1, the highest frequency is smaller in this case. For example when *N=10*, *T=500*, we estimate three factor 85% of the times in Table 1, but when the SNR is reduced to 0.30 (Table 4) this frequency decreases to 75%. Note that in all the cases the highest frequency is higher that 60%. When the SNR of SMB is reduced to 0.20 (Table 4 panel B), the correct number of factors is not always estimated with the highest frequency. For example in the case of *N=12 and T=250* we estimate 2 factors 56% of the time and 3 factors 43% of the time. This confirms that using the HFE in the presence of weak factors can lead to underestimation of the true number of factors. Note that in this case, the second highest frequency is always bigger that 25%.

While it is somewhat arbitrary, given these results, we define the factor with SNR smaller than or equal to 0.20 as a "weak" factor. Our simulation results show that when a weak factor is present in data, our HFE estimator underestimates the true number of factors. That is, the HFE estimator has only limited power to detect the factors with SNR smaller than or equal to 0.20. In the case a weak factor is present in the data the highest frequency in general will be around 60% or less and the second highest frequency will be 25% or more using BIC2. The true number of factors can estimated by analyzing the second highest frequency of BIC2 in this case.

On order to confirm our results in this preliminary one sample analysis, as we have done 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 by the HFE. We repeat this experiment for 1,000 generated samples with two different SNRs of the second common component: 0.30 and 0.20. Results are presented in Table 5.

When SNR = 0.30 (Panel A), *BIC2* penalty function is the most accurate estimating the correct number of factors. In average *BIC2* estimates 3 factors for all combinations of *N* and *T*, with exception of *N=25 T=250*. All the other penalties and the sequential hypothesis testing method are more sensitive to the presence of low SNR since they estimate in average less than 3

[18] We just report results using model selection criterion with penalty BIC2 in order to save space. Results using any other criterions are similar and are available from the authors.

factors. This results confirm that HFE still estimates the correct number of factors for a SNR of 0.30.

When STR=0.20 our method estimates three factors less often and two factors more often as reflected in the averages presented in Table 5 Panel B. For example, using *BIC2* and STR=0.20 when $N = 12$, $T=500$ the average number of factors is 2.75 (Panel B) compared with an average of 3.00 when the SNR=0.30 (Panel A). Results using BIC and BIC3 are even more extreme, since in these cases HFE estimator predicts in average 2 factors. Thus, the HFE estimator is downward biased in the presence of weak factors. Looking to the second highest frequency when it is larger than 25% it is important in this case, since it can be viewed as an evidence of a weak factor. Bai and Ng methods are not able to identify weak factors even in the case $N=25$, $T=1000$, for example they estimate in average 2.01 using IC1 in this case.

The main results from our simulation can be summarized as follows. The HFE estimator obtained from 100 random partitions performs quite well when all factors are strong. In this case sequential hypothesis testing and model selection criterion methods with any penalty function estimate the same number of factors with almost equal frequency. The HFE method performs better that Bai and Ng (2002) method for N≤15. For the case of N>15 HFE is recommended just if the sample size is bigger than 500.

Second the HFE is robust to time series autocorrelation. Simulations show that model selection criterion with non-optimal weighting matrix has good finite sample properties. This result makes model selection method a better choice compared with sequential hypothesis testing. This is because SHT requires the use of the optimal weighting matrix which involves to optimally pick the bandwidth parameter in the Newey West weighting matrix.

Third, in presence of weak factors the HFE method may lead to underestimation. BIC penalty criterion is the most accurate in finite sample detecting factors with high signal to noise ratio (strong factors). BIC2 tends to capture weaker factors, (SNR smaller than 0.20) so results using both criterions should be compared. In the case a weak factor is present in the data the highest frequency in general will be around 60% or less and the second highest frequency will be 25% or more using BIC2. The second higher frequency of BIC2 (higher than 25%) should be analyzed in case results from BIC and BIC2 are significantly different. That is, if the estimated number with the second highest frequency is greater than the HFE estimate, then that number is likely to be the correct number of factors.

6. EMPIRICAL APLICATION

We use the developed methodology to analyze the international stock return comovements. In the last years markets have become more integrated at a world level through increased capital and trade integration. Linear factors models are a generalized methodology to analyze comovements in international stock markets, given their parsimony and practical nature. Some examples are the world CAPM, the Fama-French 1998 three factor model, the world APT and the Heston and Rouwenhorst (1994) model. The main differences between these models are how many factors are included, and how the factors are estimated or constructed.

Our empirical application has two main objectives. First, to estimate the number factors that explains the international stock market. Second, in an effort to test market integration, we want to measure the level of correlation between US stock market factors and the world latent factors. Both questions can be answered using the GMM methodology proposed in this paper.

A detailed and careful study of the performance of different factor models in international stock returns can be found in Bekaerta, Hodricka and Zhang 2005 (BHZ). Based on country portfolios they analyze the explanatory power of the different factors models. They conclude that an APT model accommodating global and local factors, best fits the covariance structure. Also a factor model that embeds both global and regional Fama-French (1998) factors comes pretty close in performance. The study of BHZ, as several others, includes only information of stock markets of developed countries. This is obvious given the availability of data and the specific research question they try to answer. The use of such data sets, leads to the inclusion of a large number of European countries in the sample, and a very few countries from the Americas and Africa. In this case any factor model estimation will exacerbate the effect of regional (European) factors.

With this motivation we propose to use data on stock markets keeping the regional proportions almost constant. That is, we try to include the same number of countries from every region. The GMM methodology proposed in this paper is very adequate in this case since few cross-sectional observations are available from regions like Africa, South America or the Middle

East. We divide the world in four regions: Americas, Europe, Asia and Africa-Middle East. Our data set includes returns of stock market indexes for these four regions as follows:

Americas:	US (S&P500), Mexico (IPC), and Brazil (Bovespa).
Europe:	UK(FTSE), Germany (DAX), and Switzerland (SMI).
Asia:	Japan (Nikkei225), Hong Kong (Hang Seng) and Korea (Kospi).
Africa-Middle East:	Israel (TASE100) and South Africa (Dow Jones Titans 30).[19]

.

We use 420 weekly excess[20] return observations of the country indexes from January 2000 to January 2008. The validity of using index data to represent the market performance individual countries can be questionable. Differences in methodologies and number of companies included in each index can lead to estimation problems. We acknowledge this fact, but since our objective is not to estimate factors or risk premiums, we believe that our index data can be safely used as a good proxy for the country portfolios. Also, given the illustrative character of our empirical application, the present study can by extended with a richer data set.

Descriptive statistics of our data are presented in Table 6. Correlations of stock returns within countries in the same region are high. Mean and variances also differ across countries and regions. Results of estimation of the number of factors are presented in Table 7. We use the model selection method and report results using the penalty criterions BIC and BIC2. As explained before, we first partition the N response variables in two randomly selected groups of $N/2$ elements. We estimate the number of factors and repeat the estimation for 100 different random partitions. We report the relative frequency (percentage) for each number of factors.

Panel A shows estimation results including all 11 countries in or sample. BIC predicts 1 factor 92% of the time. BIC2 predicts 2 factors 36% of the time and 3 factors 49% of the time. As we showed in simulations BIC2 is able to capture weaker factors, and BIC is the most consistent identifying strong factors. We can conclude then that one strong factor explains the international stock market returns. There is also evidence of two weak factors in sense that they just explain an small fraction of the total variation of the response variable.

[19] All data comes from Yahoo Finance except Dow Jones Titans 30 obtained from Dow Jones web site.
[20] The risk free interest rate is the one month Treasury bill rate from K French online data library.

The weak factors can also be interpreted as local or region specific factors. In order to further analyze this possibility we estimate the number of factors for different sub-sets of the response variable. The intuition for the following analyses is that, if we remove some region or country from the sample and the weak factor becomes weaker (i.e. it is estimated with lower frequency), then this weak factor is highly related with the excluded region. Similarly if after we exclude some region from the sample, the weak factor becomes stronger (in the sense that is most frequently estimated), then this factor is not related with the excluded region.

Panel B includes results excluding US stock market index from the sample. Results are very similar compared with Panel A, so it appears that the weak factors are not specific to US market. Similar experiment is presented in Panel C, but now we exclude the 3 indexes of the region Americas After removing Americas from the sample, it seems that one of the weak factors disappear, since BIC and BIC2 now estimate 2 factors with the highest frequency. This implies that this weak factor is related to the Americas market. Similarly when we exclude Europe (Panel D) one of the weak factors totally disappear, since we do not estimate 3 factors under any criterion. This is clear evidence that one the weak factors is mostly a European factor. Panel E include results excluding Asia, results are not much different compared with Panel A. Finally when remove Africa from the sample (Panel F) we estimate three factors more often, which implies that the weak factor is not related with Africa.

Summarizing we confirm the presence of one strong global factor. Also there is evidence of a two weak factors one related mostly with European markets and the other with the Americas markets

The second part of our empirical study analyzes the relation of US market factors with the global factors. Under market integration it is intuitive to think that US market factors can be correlated in some degree with the global factors. We do not attempt to estimate neither the number of factors nor the factors for the US stock market. We rather use the most widely used proxy for the US stock market factors, the Fama-French benchmark factors. Descriptive statistics for weekly observations of EMR (excess market return), SMB, and HML are presented in Table 6[21]. As explained in section 4.2 if we use observable instruments (in this case EMR , SMB, and HML for US) in our GMM moment conditions, will be able to test how many of the global unobservable factors are correlated with our instruments. From the results presented in Table 7,

[21] Data for Fama-French factors comes from K. French online data library.

Panel G, we can conclude that the US Fama-French factors are highly correlated with the strong global factor. We estimate 71% of the time one factor using BIC1 and 74% using BIC2. For robustness we remove US from the sample and repeat the estimation. Results in Panel H are consistent with the conclusion that Fama-French factor is correlated with the strong global factor.

7. CONCLUDING REMARKS

In this paper we present a method to consistently estimate the number of factors in a linear factor models. The test is independent of the factors, since it is assumed that they are unobservable. The use GMM allows to control for time series autocorrelation and heteroskedasticity and cross sectional heteroskedasticity of the idiosyncratic errors. The estimation procedure is designed for panel data sets where the number of cross sectional observations is small and the number of time series ones is large, or vice versa. Simulation results show that our method is preferred to Bai and Ng (2002) method when one dimension of the data (N or T) is smaller of equal than 15.

Since the method requires the partition of the response variable in two groups, we propose that the estimation should be repeated 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. The estimator we propose is the "highest frequency estimator" (HFE), which is simply the number of factors most often estimated from 100 different partitions. Each partition should include around half of the elements of the response variable (N/2).

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 specially in the case of autocorrelated errors. We recommend use of the model selection method since it does not require bandwidth selection or adjustment of significance levels. BIC penalty criterion is the most accurate in finite sample detecting factors with high signal to noise ratio (strong factors). BIC2 tends to capture weaker factors, (SNR smaller than 0.20) so results using both criterions should be compared. The second higher frequency (higher that 25%) should be analyzed in case results from BIC and BIC2 are significantly different.

As an empirical application we apply our methods to the international stock market. Our empirical results imply that the returns on international stock markets are determined by one strong global factor. There is also evidence for the presence of two weak factors, one mostly related with European markets and the other with stock markets of the Americas. Finally, we conclude that the factors explaining the US stock market are highly correlated with the strong global factor. This can suggest some kind of world market integration.

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, 2008, 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.

Bekaert, Geert , Robert J. Hodrick and Xiaoyan Zhang 2005, International Stock Return Comovements. *Johnson School Research paper series No. 35-06*

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.

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 W and $\tilde{W}$. 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 $G = (P - L^*)(Q - L^*)$, $\lambda_1,...,\lambda_G$ are the *positive* eigenvalues of A, and the ξ_j are independent $\chi^2(1)$ random variables. Observe that G is the degrees of over identifying 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),

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

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 $\Pr(\hat{L} > L_o) \rightarrow 0$ and $\Pr(\hat{L} < L_o) \rightarrow 0$, as $T \rightarrow \infty$. We first show $\lim_{N\rightarrow\infty} \Pr(\hat{L} > L_o) = 0$. 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 $\Pr(\hat{L} > L_o) \leq \Pr[MS_T(L_o) - MS_T(L_a) > 0]$. Note also that

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

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

We now show that $\lim_{T\rightarrow\infty} \Pr(\hat{L} < L_o) = 0$. 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, $\Pr(\hat{L} < L_o) \to 0$.

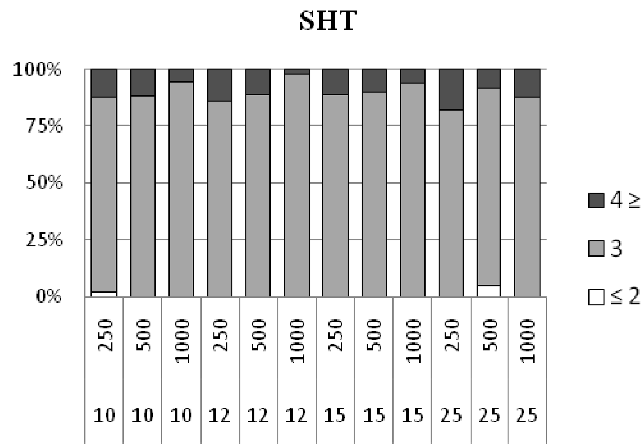
TABLE 1

Effect of Changing Partitions

A *single* data set is generated from a three-factor model. The number of factors is estimated by the sequential hypothesis testing (SHT) in panel A and model selection criterion (MSC) in panel B. For MSC we use the BIC2 penalty function. The estimation is repeated for 100 randomly chosen partitions of the generated response variables. The tables and figures below show the percentage of times each number of factors is estimated in 100 partitions for different combinations of N and T. Idiosyncratic errors are heteroskedastic but serially independent.

Panel A: Sequential Hypothesis Testing

N	T	≤ 1	2	3	4	≥ 5
10	250	0.0	2.4	83.9	12.3	-
10	500	0.0	0.0	84.5	11.5	-
10	1000	0.0	0.0	96.0	6.0	-
12	250	0.0	0.0	85.0	14.0	0.0
12	500	0.0	0.0	89.0	11.0	1.0
12	1000	0.0	0.0	98.0	2.0	0.0
15	250	0.0	0.0	88.0	11.0	1.0
15	500	0.0	0.0	90.0	10.0	0.0
15	1000	0.0	0.0	92.0	6.0	2.0
25	250	0.0	0.0	81.0	18.0	1.0
25	500	0.0	5.0	85.0	8.0	1.0
25	1000	0.0	0.0	87.0	12.0	1.0



Panel B: Model Selection Criterion

N	T	≤ 1	2	3	4	≥ 5
10	250	0.0	2.2	80.3	17.5	-
10	500	0.0	0.0	85.0	15.0	-
10	1000	0.0	0.0	94.0	6.0	-
12	250	0.0	0.0	84.0	15.0	1.0
12	500	0.0	0.0	86.0	13.0	1.0
12	1000	0.0	1.0	94.0	5.0	1.0
15	250	0.0	0.0	87.0	12.0	1.0
15	500	0.0	0.0	94.0	5.0	1.0
15	1000	0.0	0.0	95.0	5.0	0.0
25	250	83.0	1.0	16.0	0.0	0.0
25	500	0.0	0.0	85.0	15.0	0.0
25	1000	0.0	0.0	90.0	10.0	0.0

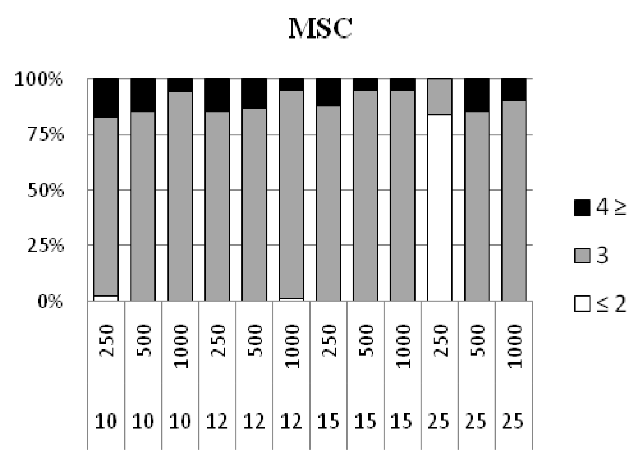


TABLE 2

Estimation of the Number of Factors by the Highest Frequency

For Panel A, 1,000 random samples are generated from a three-factor model replicating Fama-French (1993) signal to noise ratios. The number of factors is estimated by applying the sequential hypothesis testing and model selection criterion methods (MSC) to 100 randomly chosen partitions of response variables. The estimated number of factors for each sample is the number estimated the most frequently (HFE). We report the average estimated number of factors for the 1000 simulations for different combinations of *N* and *T*. We report estimation results for SHT and for MSC using penalty functions BIC, BIC2, BIC3 and AIC. We also report estimation results using Bai-Ng 2002 methods with penalty functions IC1, IC2, IC3, PC1, PC2 and PC3. Panel B includes similar results when data is generated using just one factor (SMB).

Panel A: Data Generated with 3 factors

N	T	SHT	BIC	BIC2	BIC3	AIC	IC1	IC2	IC3	PC1	PC2	PC3
10	250	3.00	2.98	3.01	2.95	3.04	8.00	8.00	8.00	8.00	8.00	8.00
10	500	3.00	3.00	3.00	3.00	3.04	8.00	8.00	8.00	8.00	8.00	8.00
10	1000	3.00	3.00	3.00	3.00	3.03	8.00	8.00	8.00	8.00	8.00	8.00
12	250	3.00	2.98	3.00	2.99	3.02	7.91	7.77	8.00	8.00	8.00	8.00
12	500	3.00	3.00	3.00	3.00	3.02	7.72	7.50	7.92	8.00	8.00	8.00
12	1000	3.00	3.00	3.00	3.00	3.01	7.36	7.15	7.60	8.00	8.00	8.00
15	250	3.00	1.05	3.00	3.00	3.00	3.16	3.10	3.28	7.86	7.72	7.98
15	500	3.00	3.00	3.00	3.00	3.00	3.09	3.07	3.13	7.62	7.49	7.79
15	1000	3.00	3.00	3.00	3.00	3.01	3.04	3.03	3.06	7.35	7.27	7.47
25	250	3.00	1.00	1.06	1.00	2.63	3.00	3.00	3.00	4.24	4.01	4.75
25	500	3.00	1.00	3.00	3.00	3.00	3.00	3.00	3.00	3.82	3.69	4.16
25	1000	3.01	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.53	3.45	3.71

Panel B: Data Generated with 1 factor

N	T	SHT	BIC	BIC2	BIC3	AIC	IC1	IC2	IC3	PC1	PC2	PC3
6	250	1.02	1.00	1.06	1.00	1.11	6.00	6.00	6.00	6.00	6.00	6.00
6	500	1.02	1.00	1.03	1.00	1.11	6.00	6.00	6.00	6.00	6.00	6.00
6	1000	1.01	1.00	1.01	1.00	1.09	6.00	6.00	6.00	6.00	6.00	6.00
8	250	1.01	1.00	1.01	1.00	1.05	8.00	8.00	8.00	8.00	8.00	8.00
8	500	1.02	1.00	1.00	1.00	1.05	8.00	8.00	8.00	8.00	8.00	8.00
8	1000	1.01	1.00	1.00	1.00	1.03	8.00	8.00	8.00	8.00	8.00	8.00
10	250	1.01	1.00	1.00	1.00	1.01	8.00	8.00	8.00	8.00	8.00	8.00
10	500	1.01	1.00	1.00	1.00	1.02	8.00	8.00	8.00	8.00	8.00	8.00
10	1000	1.01	1.00	1.00	1.00	1.02	8.00	8.00	8.00	8.00	8.00	8.00
12	250	1.01	1.00	1.00	1.00	1.01	1.49	1.24	2.87	8.00	8.00	8.00
12	500	1.01	1.00	1.00	1.00	1.01	1.22	1.17	1.29	8.00	8.00	8.00
12	1000	1.01	1.00	1.00	1.00	1.00	1.14	1.12	1.18	8.00	8.00	8.00
15	250	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.04	7.73	7.50	7.96
15	500	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	7.28	7.14	7.55
15	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	6.98	6.87	7.12
25	250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3.10	2.84	3.69
25	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.66	2.52	2.98
25	1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.39	2.29	2.56

TABLE 3

Effects of Autocorrelation

1,000 random samples are generated from a three-factor model replicating Fama-French (1993) signal to noise ratios. Idiosyncratic errors follow an AR(1) process with ρ=0.5. The number of factors is estimated by applying the sequential hypothesis testing and model selection criterion methods (MSC) to 100 randomly chosen partitions of response variables. The estimated number of factors for each sample is the number estimated the most frequently (HFE). We report the average estimated number of factors for the 1000 simulations for different combinations of *N* and *T*. We report estimation results for SHT and for MSC. For SHT the weighting matrix is computed using Newey West methods with bandwidth 0 and 3.For MSC the weighting matrix is White and penalty functions BIC, BIC2, BIC3 and AIC. We also report estimation results using Bai-Ng 2002 methods with penalty functions IC1, IC2, IC3, PC1, PC2 and PC3. Panel B includes similar results when data is generated using just one factor (SMB).

		SHT		MSC				BAI-NG					
N	**T**	**BW=0**	**BW=3**	**BIC**	**BIC2**	**BIC3**	**AIC**	**IC1**	**IC2**	**IC3**	**PC1**	**PC2**	**PC3**
10	250	3.10	3.00	2.98	3.23	2.95	3.40	8.00	8.00	8.00	8.00	8.00	8.00
10	500	3.09	3.01	3.00	3.12	3.00	3.37	8.00	8.00	8.00	8.00	8.00	8.00
10	1000	3.08	3.01	3.00	3.07	3.00	3.33	8.00	8.00	8.00	8.00	8.00	8.00
12	250	3.22	3.01	2.95	3.21	2.98	3.47	8.00	7.96	8.00	8.00	8.00	8.00
12	500	3.24	3.01	3.00	3.11	3.00	3.50	7.94	7.89	7.98	8.00	8.00	8.00
12	1000	3.21	3.01	3.00	3.03	3.00	3.45	7.71	7.62	7.81	8.00	8.00	8.00
15	250	3.55	1.00	1.00	3.09	2.93	3.48	3.36	3.25	3.76	7.98	7.94	8.00
15	500	3.56	3.01	3.00	3.04	3.00	3.57	3.15	3.13	3.22	7.90	7.83	7.96
15	1000	3.57	3.02	3.00	3.01	3.00	3.59	3.08	3.07	3.10	7.60	7.52	7.68
25	250	4.92	1.00	1.00	1.00	1.00	1.00	3.00	3.00	3.01	5.04	4.81	5.62
25	500	5.70	1.00	1.00	3.00	2.74	3.03	3.00	3.00	3.00	4.31	4.17	4.60
25	1000	5.75	3.00	3.00	3.00	3.00	3.32	3.00	3.00	3.00	3.85	3.79	4.00

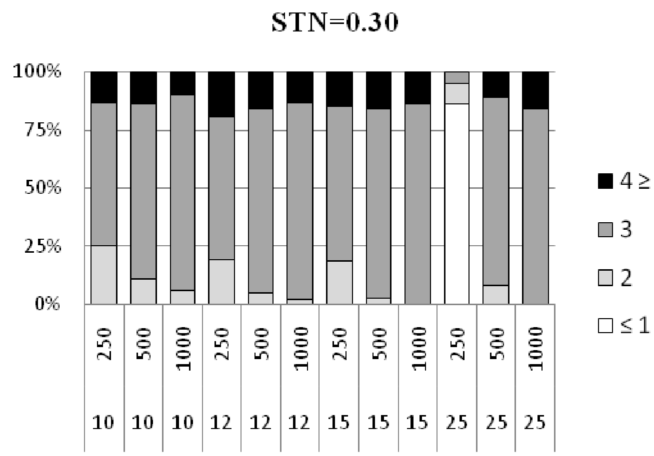
TABLE 4

Effect of Changing Partitions with a Weak Factor

A *single* data set is generated from a three-factor model replicating Fama-French (1993) signal to noise ratios for factors EMR and HML. The signal to noise ratio of factor SML is reduced to 0.30 for Panel A and to 0.20 for Panel B. The number of factors is estimated by the model selection criterion (MSC) method using the penalty function BIC2. The method is applied to 100 randomly chosen partitions of response variables. The tables and figures show the percentage of times each number of factors is estimated for different combinations of *N* and *T*.

Panel A: Signal to noise 0.30

N	T	≤ 1	2	3	4 ≥
10	250	0.0	25.0	62.0	13.0
10	500	0.0	11.0	75.0	14.0
10	1000	0.0	6.0	84.0	10.0
12	250	0.0	19.0	62.0	19.0
12	500	0.0	5.0	78.0	16.0
12	1000	0.0	2.0	85.0	13.0
15	250	0.0	19.0	67.0	15.0
15	500	0.0	3.0	81.0	16.0
15	1000	0.0	0.0	86.0	14.0
25	250	86.0	9.0	5.0	0.0
25	500	0.0	8.0	81.0	11.0
25	1000	0.0	0.0	84.0	16.0



Panel B: Signal to noise 0.20

N	T	≤ 1	2	3	4 ≥
10	250	0.0	60.0	35.0	5.0
10	500	0.0	42.0	50.0	8.0
10	1000	0.0	23.0	67.0	10.0
12	250	0.0	59.0	33.0	8.0
12	500	0.0	33.0	56.0	10.0
12	1000	0.0	14.0	72.0	14.0
15	250	0.0	66.0	30.0	4.0
15	500	0.0	33.0	56.0	10.0
15	1000	0.0	8.0	77.0	15.0
25	250	71.0	28.0	1.0	0.0
25	500	0.0	50.0	48.0	2.0
25	1000	0.0	4.0	77.0	19.0

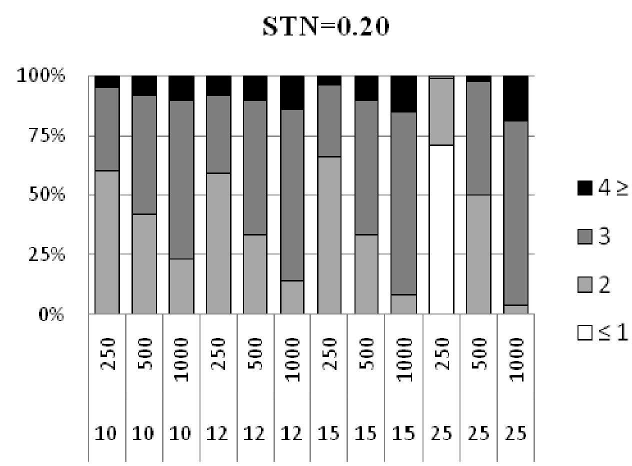


Table 5

Highest Frequency Estimation of the Number of Factors under Weak Factors

1,000 random samples are generated from a three-factor model replicating Fama-French (1993) signal to noise ratios for factors EMR and HML. The signal to noise ratio of factor SMB is reduced to 0.30 for Panel A and to 0.15 for Panel B The number of factors is estimated by applying the sequential hypothesis testing and model selection criterion methods (MSC) to 100 randomly chosen partitions of response variables. The estimated number of factors for each sample is the number estimated the most frequently (HFE). We report the average estimated number of factors for the 1000 simulations for different combinations of *N* and *T*. We report estimation results for SHT and for MSC using penalty functions BIC, BIC2, BIC3 and AIC. We also report estimation results using Bai-Ng 2002 methods with penalty functions IC1, IC2, IC3, PC1, PC2 and PC3.

Panel A: Signal to Noise SMB= 0.30

N	T	SHT	BIC	BIC2	BIC3	AIC	IC1	IC2	IC3	PC1	PC2	PC3
10	250	2.86	2.12	2.88	2.03	2.98	8.00	8.00	8.00	8.00	8.00	8.00
10	500	2.98	2.66	2.98	2.45	3.06	8.00	8.00	8.00	8.00	8.00	8.00
10	1000	3.00	2.95	3.00	2.89	3.05	8.00	8.00	8.00	8.00	8.00	8.00
12	250	2.95	2.04	2.94	2.04	3.00	7.88	7.61	7.99	8.00	8.00	8.00
12	500	3.00	2.76	3.00	2.76	3.03	7.63	7.24	7.87	8.00	8.00	8.00
12	1000	3.01	2.98	3.00	2.98	3.03	7.12	6.92	7.42	8.00	8.00	8.00
15	250	2.99	1.16	2.96	2.08	3.00	3.06	2.99	3.19	7.85	7.71	7.98
15	500	3.01	2.70	3.00	2.92	3.01	3.02	2.99	3.08	7.61	7.48	7.79
15	1000	3.01	2.99	3.00	3.00	3.01	3.00	2.98	3.02	7.34	7.26	7.46
25	250	2.63	1.00	1.01	1.00	1.71	2.99	2.97	3.00	4.20	3.98	4.74
25	500	3.01	1.00	3.00	2.78	3.00	2.99	2.99	3.00	3.81	3.66	4.13
25	1000	3.01	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.52	3.43	3.69

Panel A: Signal to Noise SMB= 0.20

N	T	SHT	BIC	BIC2	BIC3	AIC	IC1	IC2	IC3	PC1	PC2	PC3
10	250	2.23	2.00	2.25	2.00	2.46	8.00	8.00	8.00	8.00	8.00	8.00
10	500	2.66	2.01	2.59	2.00	2.87	8.00	8.00	8.00	8.00	8.00	8.00
10	1000	2.95	2.25	2.89	2.11	3.03	8.00	8.00	8.00	8.00	8.00	8.00
12	250	2.34	2.00	2.25	2.00	2.51	7.61	6.90	7.97	8.00	8.00	8.00
12	500	2.86	2.01	2.75	2.01	2.96	6.84	6.16	7.58	8.00	8.00	8.00
12	1000	2.99	2.37	2.97	2.36	3.04	5.96	5.56	6.49	8.00	8.00	8.00
15	250	2.40	1.68	2.19	2.00	2.48	2.30	2.23	2.50	7.85	7.71	7.98
15	500	2.94	2.00	2.79	2.04	2.97	2.22	2.18	2.29	7.59	7.45	7.79
15	1000	3.02	2.40	2.99	2.76	3.03	2.18	2.17	2.21	7.32	7.25	7.45
25	250	2.04	1.00	1.17	1.00	1.81	2.10	2.06	2.27	4.15	3.93	4.68
25	500	3.00	1.00	2.48	2.00	2.96	2.06	2.05	2.14	3.76	3.62	4.08
25	1000	3.04	2.04	3.00	2.92	3.00	2.05	2.04	2.07	3.48	3.41	3.65

TABLE 6

Data Description

Descriptive statistics of weekly returns stock market indexes of 11 countries for the period of January 2000 to January 2008 are presented below. Also we report descriptive statistics of weekly Fama- French benchmark factors for the same period. The bottom table includes correlation coefficients of for the excess returns of the stock market indexes.

COUNTRY	MEAN	STD. DEV	MIN	MAX
US	0.007	2.284	-11.601	7.780
MEXICO	0.372	3.109	-16.236	11.321
BRAZIL	0.382	4.026	-18.123	11.793
JAPAN	-0.030	2.770	-10.678	9.932
KOREA	0.212	3.851	-13.838	15.106
HONG KONG	0.161	2.995	-10.072	12.433
UK	-0.002	2.129	-8.482	10.594
GERMANY	0.053	3.211	-12.994	13.755
SWITZERLAND	0.038	2.482	-11.212	17.690
ISRAEL	0.217	2.793	-14.292	12.052
SOUTH AFRICA	0.312	2.703	-11.515	7.482
MKT-RF	-0.001	2.380	-13.740	9.180
HML	0.084	1.472	-9.370	6.340
SMB	0.189	1.448	-6.810	9.740

Correlations

	US	MEX	BRA	JAP	KOR	HK	UK	GER	SW	ISR	SA
US	1.00										
MEX	0.66	1.00									
BRA	0.49	0.53	1.00								
JAP	0.38	0.34	0.34	1.00							
KOR	0.43	0.45	0.41	0.56	1.00						
HK	0.47	0.47	0.38	0.50	0.60	1.00					
UK	0.75	0.58	0.44	0.41	0.40	0.55	1.00				
GER	0.74	0.55	0.49	0.45	0.46	0.53	0.80	1.00			
SW	0.68	0.44	0.33	0.37	0.35	0.46	0.78	0.76	1.00		
ISR	0.27	0.29	0.26	0.27	0.36	0.34	0.27	0.35	0.22	1.00	
SA	0.50	0.50	0.38	0.37	0.45	0.46	0.47	0.44	0.40	0.27	1.00

TABLE 7

Empirical Application

The model selection method is used to estimate the number of factors in the excess returns of country's stock market indexes. The weighting matrix used is White (1980). The estimation is repeated for 100 different partitions of the response variable. We report the percentage estimated of each factor for each specification.

Panel A: ALL COUNTRIES

# Factors	BIC	BIC2
0	6.0	0.0
1	92.0	6.0
2	2.0	36.0
3	0.0	49.0
≥4	0.0	9.0
Average	0.96	2.61

Panel B: NO US

# Factors	BIC	BIC2
0	4.0	0.0
1	81.0	9.0
2	15.0	32.0
3	0.0	45.0
≥4	0.0	14.0
Average	1.11	2.64

Panel C: NO AMERICAS

# Factors	BIC	BIC2
0	0.0	0.0
1	28.0	4.0
2	68.0	76.0
≥3	4.0	20.0
Average	1.76	2.16

Panel D: NO EUROPE

# Factors	BIC	BIC2
0	0.0	0.0
1	83.0	25.0
2	17.0	73.0
≥3	0.0	2.0
Average	1.17	1.77

Panel E: NO ASIA

# Factors	BIC	BIC2
0	0.0	0.0
1	91.0	14.0
2	9.0	50.0
≥3	0.0	36.0
Average	1.09	2.22

Panel F: NO AFRICA

# Factors	BIC	BIC2
0	0.0	0.0
1	67.0	2.0
2	25.0	25.0
≥3	8.0	73.0
Average	1.41	2.71

Panel G: FAMA FRENCH FACTORS

# Factors	BIC	BIC2
0	29.0	0.0
1	71.0	74.0
2	0.0	26.0
≥3	0.0	0.0
Average	0.71	1.26

Panel H: FAMA FRENCH NO US

# Factors	BIC	BIC2
0	8.0	0.0
1	92.0	56.0
2	0.0	44.0
≥3	0.0	0.0
Average	0.92	1.44