Journal of
Business &
Economic
Statistics

Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets

George Kapetanios

To cite this article: George Kapetanios (2010) A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets, Journal of Business & Economic Statistics, 28:3, 397-409, DOI: 10.1198/jbes.2009.07239

To link to this article: https://doi.org/10.1198/jbes.2009.07239

Published online: 01 Jan 2012.

A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets

George KAPETANIOS
Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS, U.K.
(*G.Kapetanios@qmul.ac.uk*)

The paradigm of a factor model is very appealing and has been used extensively in economic analyses. Underlying the factor model is the idea that a large number of economic variables can be adequately modeled by a small number of indicator variables. Throughout this extensive research activity on large dimensional factor models a major preoccupation has been the development of tools for determining the number of factors needed for modeling. This article provides an alternative method to information criteria as a tool for estimating the number of factors in large dimensional factor models. The new method is robust to considerable cross-sectional and temporal dependence. The theoretical properties of the method are explored and an extensive Monte Carlo study is undertaken. Results are favorable for the new method and suggest that it is a reasonable alternative to existing methods.

KEY WORDS: Factor models; Large sample covariance matrix; Maximum eigenvalue.

1. INTRODUCTION

The paradigm of a factor model is very appealing and has been used extensively in economic analyses. Underlying the factor model is the idea that a large number of economic variables can be adequately modeled by a small number of indicator variables. Factor analysis has been used fruitfully to model, among other cases, asset returns, macroeconomic aggregates, and Engel curves (see, e.g., Stock and Watson 1989; Lewbel 1991, and others).

Most analyses have traditionally focused on small datasets, meaning that the number of variables, N, to be modeled via a factor model is finite. Recently, Stock and Watson (2002) put forward the case for analyzing large datasets via factor analysis, where N is allowed to tend to infinity. Stock and Watson (2002) suggested the use of principal components for estimating factors in this context. Similar work in a more general setting has been carried out by, e.g., Forni et al. (2000, 2004) in which use of dynamic principal components has been made.

Throughout this extensive research activity on large dimensional factor models a major preoccupation has been the development of tools for determining the number of factors needed for modeling. The main tool used in econometrics for estimating the number of factors for large dimensional datasets is the use of information criteria developed by Bai and Ng (2002). The criteria developed are modifications of standard information criteria such as Akaike's information criterion where the penalty terms needed for consistent estimation of the number of factors depend both on the number of observations T as well as N, unlike the traditional criteria where the penalty terms depend only on T. Further recent work has tried to extend the work of Bai and Ng (2002). Bai and Ng (2007) and Amengual and Watson (2007) extended the work of Bai and Ng (2002) which is based on static factor model to a restricted dynamic case. Hallin and Liska (2007) provided extensions for the general dynamic case. Jacobs and Otter (2006) provided an alternative solution to the problem using a minimum entropy approach.

This article aims to provide an alternative to information criteria as tools for estimating the number of factors in large dimensional factor models. The main reason for proposing this alternative method is that Monte Carlo evidence suggests that it can be a more robust method than information criteria in determining the number of factors. Further, the approach is based on random matrix theory which, although widely used in the statistical and physics literature, is not well known in econometrics.

Previous work by Kapetanios (2004) made use of random matrix theory (RMT) to devise methods for determining the number of factors in large datasets. That work was followed up in Onatski (2005). However, a number of problems existed with the approach suggested in Kapetanios (2004). The main problem relates to the stringency of the assumptions made to derive formal results for the method. In the current article we relax most such assumptions. Further, the current method is, in fact, based on a sequence of tests on the largest eigenvalues of the sample covariance matrix. Given the available results from RMT, which will be briefly presented in the following section, it may appear surprising that we can propose an operational method based on asymptotic distributions of eigenvalues. However, we are able to do this because we use subsampling which is a resampling technique similar to the bootstrap, but much more widely applicable. The need for a resampling method is clear. Asymptotic distributional results exist only for very special cases such as the case of iid data. It is further likely that deviations from such restrictive assumptions will not only lead to different distributions, but different convergence rates too. In such an environment subsampling is likely to be the only technique available for distributional analysis for some time.

The article is organized as follows: Section 2 surveys the available results on the behavior of the eigenvalues of large

© 2010 American Statistical Association
Journal of Business & Economic Statistics
July 2010, Vol. 28, No. 3
DOI: 10.1198/jbes.2009.07239

sample covariance matrices and introduces the new method. Section 3 discusses the new method and provides some theoretical results. Results from a Monte Carlo study are presented in Section 4. An empirical application to S&P500 data is presented in Section 5. Finally, Section 6 concludes.

2. PRELIMINARIES

The factor model we consider for a given dataset for cross sectional unit i at time t, is given by

$$y_{i,t} = f_t' \lambda_i + \epsilon_{i,t}, \tag{1}$$

where f_t is an r-dimensional vector of factors at time t, λ_i is an r-dimensional vector of factor loadings for cross sectional unit i, and $\epsilon_{i,t}$ is the idiosyncratic part of $y_{i,t}$. Usually factors are assumed to be weakly dependent time series processes and the factor loadings are assumed to be random variables. We will also assume that, in general, the idiosyncratic terms are weakly dependent processes as well as with mild cross-sectional dependence. The nature of this dependence will be made clear later.

Rewriting the above model in matrix notation gives

$$Y = F\Lambda + \epsilon, \tag{2}$$

where $Y = (y_1, \dots, y_N)$, $F = (F_1, \dots, F_r)$, $\Lambda = (\lambda_1, \dots, \lambda_N)$, $\epsilon = (\epsilon_1, \dots, \epsilon_N)$, $y_i = (y_{i,1}, \dots, y_{i,T})'$, $F_j = (f_{j,1}, \dots, f_{j,T})'$, and $\epsilon_i = (\epsilon_{i,1}, \dots, \epsilon_{i,T})'$. Following Chamberlain and Rothschild (1983) and assuming uncorrelatedness between the factors and the idiosyncratic components $\epsilon_{i,t}$, it is easy to see that the covariance matrix of the dataset is given by

$$\Sigma_Y = \Sigma_f + \Sigma_\epsilon, \tag{3}$$

where Σ_f is a matrix with finite rank r and Σ_ϵ is the covariance matrix of the idiosyncratic component which is assumed to have bounded eigenvalues for all N. Under certain conditions on the factor loadings, detailed in the next section, the largest r eigenvalues of Σ_f will tend to infinity at rate N whereas the rest will be equal to zero.

Before outlining in intuitive terms the new methodology, we quote some results on large dimensional covariance matrices. Let $\epsilon = [\epsilon_{i,t}]$ denote a $T \times N$ matrix of iid mean zero and unit variance random variables. Let $\hat{\Sigma}_\epsilon$ denote the sample covariance matrix given by $\frac{1}{T}\epsilon'\epsilon$. Then, the largest eigenvalue of $\hat{\Sigma}_\epsilon$, denoted $\hat{\mu}_1$, converges almost surely to $(1+\sqrt{c})^2$, where $c = \lim_{N,T\to\infty} \frac{N}{T}$. The result is remarkable in its simplicity. For example, for $N = T$ the largest eigenvalue converges almost surely to 4. This result has been repeatedly proven under successively weaker conditions culminating in the work of Yin, Bai, and Krishnaiah (1988) who proved the result showing that a necessary and sufficient condition is that $E(\epsilon_{i,t}^4) < \infty$. In this context it has also been shown that the minimum eigenvalue of $\hat{\Sigma}_\epsilon$ converges almost surely to $(1-\sqrt{c})^2$ as long as $N < T$ and, obviously, zero otherwise. We note that the condition $E(\epsilon_{i,t}^4) < \infty$ is crucial. If this condition does not hold the maximum eigenvalue tends to infinity. It is further worth noting that the relevant literature has not addressed the case in which the sequence N/T does not converge (i.e., when $\liminf N/T \neq \limsup N/T$).

The result has been extended to more complicated setups. To appreciate the following result we note that in the case of large dimensional matrices, where the dimension of the matrix tends to infinity, focus has been placed on the limit of the empirical distribution of the eigenvalues of the matrix (referred to as empirical spectral distribution (ESD) in the literature). Thus, it has been shown, among other things, by Bai and Silverstein (1998), for a $N \times N$ nonnegative definite symmetric matrix Q_N, that the limit as $N, T \to \infty$ of the ESD of $\frac{1}{T} Q_N^{1/2} \epsilon' \epsilon Q_N^{1/2}$ has a support which is almost surely contained in the support of the limit of the empirical distribution of the eigenvalues of Q_N.

The above results relate to temporally iid data. Recently, work by Hachem, Loubaton, and Najim (2005a, 2005b) derived the limit of the ESD of the sample covariance matrix of temporally independent, but heterogeneously distributed data and temporally dependent data with absolutely summable autocovariances. In the latter case it is shown that this limit crucially depends on the coefficients of the MA representation of the data. This suggests that temporal dependence does not only affect the parameters of the asymptotic limits, but their functional form too. This necessarily implies that standard asymptotic approaches to the construction of testing procedures are likely to be of little value.

The above results deal with the form of the limits of extreme eigenvalues and the ESD. An important question concerns the rates at which these limits are approached. Unfortunately, results here are less common. The first major work to address this was Tracy and Widom (1996) who showed that the distribution function associated with the limit law of the largest eigenvalue of an $N \times N$ Gaussian symmetric matrix is given by

$$F_1(s) = \exp\left\{0.5 \int_s^\infty q(x) + (x-s)q^2(x)\,dx\right\},$$

where q solves the nonlinear Painleve II differential equation given by

$$q''(x) = xq(x) + 2q^3(x).$$

Of more relevance to our purposes is the result obtained by Johnstone (2001) who showed that

$$\frac{\hat{\mu}_1 - (1+\sqrt{c})^2}{T^{-1}(\sqrt{T-1}+\sqrt{N})(\frac{1}{\sqrt{T-1}} + \frac{1}{\sqrt{N}})^{1/3}} \Rightarrow W \sim F_1, \tag{4}$$

where $\hat{\mu}_1$ is the maximum eigenvalue of the sample covariance matrix of an $T \times N$ matrix of iid N(0, 1) variates.

Recently, there has been some further unpublished work on the distribution of the largest eigenvalues of the sample covariance matrix of temporally independent Gaussian data where the population covariances matrix is not equal to the identity, but has some eigenvalues which are different from one. This work includes Peche (2005) and Baik, Ben-Arous, and Peche (2004). The results quoted therein reinforce the fact that the relevant distributions depend in complex ways on a multitude of factors even before cross-sectional or temporal dependence is explicitly addressed. There does not seem to be any work publicly available on asymptotic distributions of the form in Equation (4) and convergence rates for temporally dependent data.

We now outline the suggested estimation method for the number of factors. It is clear that if the number of factors in the dataset is r^0 then, under some regularity conditions, the first r^0 eigenvalues of Σ_Y will increase at rate N, whereas the rest will remain bounded. The fact that the first r^0 eigenvalues of Σ_Y increase at rate N follows from Equation (3) and the fact that the r^0 largest eigenvalues of Σ_f will grow at rate N as long as the loading matrix Λ is not sparse. It is reasonable to expect a similar behavior from the eigenvalues of the sample covariance matrix. This statement will be made formal in the next section. Let us denote the eigenvalues of the sample covariance matrix of y_t by $\hat{\mu}_r$, $r = 1, \ldots, N$. Then, it is reasonable to expect that $\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}$ will tend to infinity for $r = 1, \ldots, r^0$, but remain bounded for $r = r^0 + 1, \ldots, r^{\max}$, where $r^{\max}$ is some finite number such that $r^0 < r^{\max}$. To see this, note that when there are r^0 factors in the dataset, then $\hat{\mu}_r$, $r = 1, \ldots, r^0$, increase at rate N, whereas $\hat{\mu}_r$, $r = r^0 + 1, \ldots, N$, remain bounded. Possible choices for $r^{\max}$ will be discussed in the next section. The role of $\hat{\mu}_{r^{\max}+1}$ is as an estimator of the upper bound for the maximum eigenvalue when there is no factor structure in the dataset. For example, for iid data this bound is known and equal to $(1 + \sqrt{c})^2$. In general, however, this bound is not known. However, under certain conditions on the limit of the ESD, which will be spelled out in the next section, for any finite number, ℓ, the ℓ largest eigenvalues will tend to the upper bound of the ESD and so $\hat{\mu}_{r^{\max}+1}$ is a consistent estimator for that bound.

If there is no factor structure then $\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}$, $r = 1, \ldots, r^{\max}$, suitably normalized by some sequence of constants depending on N and T, denoted $\tau_{N,T}$, should converge to some limit law. In the presence of factors it should tend to infinity at rate $N\tau_{N,T}$. If the limit law and $\tau_{N,T}$ are known then the null hypothesis that the true number of factors, r^0 in the dataset, is equal to r ($H_{0,r} : r^0 = r$) against the alternative hypothesis $H_{1,r} : r^0 > r$ could be tested by considering the test statistic $\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}$. Unfortunately this limit law is not known and given the available results discussed above it is highly likely to depend in complicated ways on the characteristics of the data such as temporal and cross-sectional dependence. As will be argued in the next section, the form of $\tau_{N,T}$ itself is likely to depend on the ESD of the dataset. Hence, asymptotic analysis is likely to be problematic. A standard solution in such cases is to consider the bootstrap. Unfortunately, asymptotic validity of the standard bootstrap is difficult to establish as well, since necessary uniform smoothness conditions with respect to the limit law are likely to be very hard to establish. We suggest an alternative resampling technique, referred to as subsampling, which is asymptotically valid under minimal conditions. Using this technique the exact distribution of $\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}$ can be approximated and a test can be carried out. Then, a sequence of such tests can be used to determine the true number of factors in the dataset. Such an approach has a long history in econometrics and statistics for solving similar inference problems such as, e.g., the determination of the rank of matrices from their estimated counterparts. In particular, the problem may be thought of as one of determining the rank of Σ_f when only an estimate of Σ_Y is available. Then, it follows a considerable body of work on rank determination using a sequence of tests such as Camba-Mendez et al. (2003) and Camba-Mendez and Kapetanios (2005).

Before concluding this section it is worth reviewing the various methods of determining the number of factors that are available in the literature, including the methods that make use of information criteria. We first cover the information criteria methods. The most widely used criteria are those proposed by Bai and Ng (2002) which can be used for consistent estimation of r in Equation (1). A summary of these is given in the Monte Carlo study of this article in Section 4. Bai and Ng (2007) and Amengual and Watson (2007) extended the work of Bai and Ng (2002) by assuming that

$$f_t = \sum_{j=1}^{p} A_j f_{t-j} + \eta_t,$$

where $\eta_t = \Xi\psi_t$, Ξ is an $r \times q$ matrix with full column rank, $q \leq r$, and ψ_t is a martingale difference sequence. This setup essentially specifies that there exist q primitive shocks, ψ_t. Both Bai and Ng (2007) and Amengual and Watson (2007) provided methods for consistently estimating q. In both articles the information criteria of Bai and Ng (2002) are used to provide an estimate of r which is then used to estimate the r factors by principal components. Once these estimated factors, denoted by $\hat{f}_t$, are extracted, the two articles proceeded in two different ways to estimate q. Bai and Ng (2007) fit a VAR model on $\hat{f}_t$ and estimated the rank of the covariance matrix of the residuals, whereas Amengual and Watson (2007) constructed $y_t^* = y_t - \sum_{j=1}^{p} \Lambda A_j f_{t-j}$ and determined the number of factors in y_t^* using the information criteria of Bai and Ng (2002). Our method can be used in the context of both Bai and Ng (2007) and Amengual and Watson (2007), since it will provide an initial estimate of r and, in the case of Amengual and Watson (2007), be used in the place of the information criteria of Bai and Ng (2002) to provide the estimate of q.

Hallin and Liska (2007) further extended the model under consideration by specifying that

$$y_{i,t} = \sum_{j=1}^{q} b_{i,j}(L)\eta_{j,t} + \epsilon_{i,t},$$

where $b_{i,j}(L), j = 1, \ldots, q$, are lag polynomials. These lag polynomials can be of finite or infinite order. This is the same model used in a series of articles by Forni, Hallin, Lippi, and Reichlin, such as, e.g., Forni et al. (2000, 2004), and is referred to as the generalized dynamic factor model. If all the lag polynomials $b_{i,j}(L), j = 1, \ldots, q$, are of finite order, then this model can be written in the form of Equation (1), but, of course, in that representation the number of factors is not q, but $q(p_1 + \cdots + p_q)$, where p_j is the maximum lag order in $b_{i,j}(L)$ across i. Hallin and Liska (2007) proposed a new information criterion for determining q in this case. The new criterion uses the eigenvalues of the spectral density matrix of y_t. As the generalized dynamic factor model is more general than the simple dynamic factor model of Equation (1), it is clear that the criteria of Hallin and Liska (2007) are more widely applicable than those of Bai and Ng (2002).

Another article that addresses the issue of determining the number of factors is Onatski (2006). That article is, in some sense, the closest to our work. In particular, Onatski (2006) considered test statistics that are based on the eigenvalues of the sample covariance matrix of y_t and used these to construct tests that can be used to determine the number of factors. However, unlike our approach, asymptotic approximations to the distribution of the test statistics are derived. These derivations draw heavily on the random matrix theory which we briefly presented above. As we discuss above, application of this theory requires stringent assumptions. In particular, Onatski (2006) assumed that the data are Gaussian and temporally independent. However, these assumptions are unlikely to hold for economic data.

3. THEORY

In this section we discuss the theoretical properties of the new method. For that we impose the following set of assumptions:

Assumption 1. $T^{-1}\sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \Sigma$ as $T \to \infty$, for some $r \times r$ positive definite matrix Σ. f_t is a strictly stationary process.

Assumption 2. $E(\lambda_i \lambda_i') = D$ for some positive definite matrix D. The observed ordering of the cross-sectional units is such that the sequences $\{\epsilon_i\}_{i=1}^{N}$ and $\{\lambda_i\}_{i=1}^{N}$ are cross-sectionally strong mixing processes in the sense of Connor and Korajczyk (1993) with mixing size equal to $\zeta > 4$.

Assumption 3. $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = \sigma_i^2$, $E|\epsilon_{i,t}|^4 \leq M$ for some constant M. The maximum eigenvalue of $1/T\epsilon'\epsilon$ is $O_p(1)$ as $N, T \to \infty$. $\epsilon_{i,t}$ is strictly stationary over i and t. λ_i is strictly stationary over i. f_t and ϵ_t are independent processes.

Assumption 4. Let r^0 denote the true number of factors. Then, for every $r = r^0 + 1, \ldots, r^{\max}$, there exists a sequence of constants $\tau_{N,T}$, bounded away from zero, such that $\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}) \xrightarrow{d} J_r$ where J_r is a limit law, as $N, T \to \infty$. Further, for $r = 1, \ldots, r^0$, the maximum of B random draws from the asymptotic distribution of $\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1})$ is $O_p(B^{\kappa} N \tau_{N,T})$ for some $0 < \kappa < 1$.

Assumption 5. $N, T \to \infty$ in such a way that $N/T \to c$, where $0 \leq c < \infty$.

Remark 1. Assumptions 1 and 3 impose conditions on the factors and idiosyncratic errors. Assumption 3 imposes, in addition, boundedness of the maximum eigenvalue of $1/T\epsilon'\epsilon$. For cross-sectional dependence this follows immediately by the boundedness of the eigenvalues of Σ_ϵ which follows from Assumption 2, Lemma 1 below, and theorem 1.1 of Bai and Silverstein (1998). The assumption is, therefore, included to cover the case of temporal dependence for $\epsilon_{i,t}$. The assumption of finite fourth moments for $\epsilon_{i,t}$ is minimal. It is required even when $\epsilon_{i,t}$ are cross-sectionally and temporally independent to obtain bounded eigenvalues for $\hat{\Sigma}_\epsilon$.

Remark 2. The mixing assumption in Assumption 2 follows the pioneering work of Connor and Korajczyk (1993) on cross-sectional mixing. For more details on the more common concept of temporal mixing, see, e.g., Davidson (1994) or Doukhan (1994). Note that, as it stands, the assumption requires that the researcher has some knowledge of what generates cross-sectional dependence. In particular, requiring mixing in the observed sample rules out situations where the sample is collected in a way that "revisits" correlated units. As an example, if households in the same neighborhood are correlated they need to be sampled consecutively as a block. However, this assumption is asymptotic allowing for considerable cross-sectional correlation in the idiosyncratic component.

Remark 3. Assumption 4 is not standard. There are two parts to it: the first posits the existence of a limit law for the normalized difference of the eigenvalues. No additional assumptions are placed on that limit law. Further, nothing is assumed about the rate of convergence to that limit law. Secondly, a very mild assumption is made about the maximum of a set of draws of $\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1})$, in the case where there exists a factor structure and $r \leq r^0$. This is needed in the proof of the consistency of the estimator of the number of factors to ensure that extreme subsampled statistics are smaller in order of magnitude than the original statistic. Note that any draw of $\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1})$ for $r = 1, \ldots, r^0$ is $O_p(N\tau_{N,T})$ as $\hat{\mu}_r$ diverges. This assumption is related to extreme value statistics. A cursory scan of extreme value theory (see, e.g., Galambos 1978) suggests that all distributions, which are bounded in probability, that have been considered in this literature, satisfy the assumption that the maximum from a set of B draws from the distribution is $O_p(B^{\kappa})$ for some $0 < \kappa < 1$. In fact, for many commonly used distributions the order of magnitude is much smaller. For example, for the χ^2 it is $\ln B$. Overall, we feel that Assumption 4 is mild. It is, however, impossible to verify without knowing more about the asymptotic properties of maximum eigenvalues. Given our discussion in Section 2 such knowledge is extremely difficult to obtain using available theory.

Remark 4. Assumption 4 requires a choice for the constant $r^{\max}$. In theory this just needs to be a sufficiently large finite number. In practice, prior views about the potential factor structure of the dataset under consideration may be used. In this context it might be reasonable to make $r^{\max}$ a function of N. Our suggestion to relate $r^{\max}$ to N appears reasonable if one realizes that this choice bears some resemblance to selecting the maximum lag order in dynamic models. In that context the maximum lag order is usually intimately related to the number of observations. Another practical suggestion is to start with some value for $r^{\max}$ and if the number of factors chosen is equal to or close to that value then to increase $r^{\max}$ and estimate the number of factors again. We note that this issue is not very easy to address and has, therefore, received little attention in the existing literature.

Assumption 2 is related to the more usual conditions on the maximum eigenvalue of Σ_ϵ by Lemma 1 below which provides a lower bound for the mixing size of ϵ_i.

Lemma 1. The maximum eigenvalue of Σ_ϵ is bounded if $\delta > 2/(\zeta - 2)$, where $E(|\epsilon_{i,t}|^{2+2\delta}) < \infty$ and ζ is the mixing size.

The lemma was proven in Connor and Korajczyk (1993).

We next discuss the subsampling methodology for estimating J_r. Subsampling was introduced informally by Mahalanobis (1946). Its properties were first discussed formally in Politis and Romano (1994). The method entails resampling without replacement from the original data and constructing samples of smaller size than the original sample. By virtue of the fact that, as Politis, Romano, and Wolf (1999, p. 40) put it, "each subset of size b (taken without replacement from the original data) is indeed a sample of size b from the true model," a more robust approximation to the properties of statistics based on the original sample is feasible. In our case, we need to address the fact that the data are both cross-sectionally and temporally dependent. In these cases block resampling is suggested by Politis and Romano (1994). Therefore, we consider block resampling that retains both the temporal and cross-sectional order of the original sample. In our case we need to resample blocks in two dimensions. In this way the original orderings are retained and the resample is indeed a sample of size b from the true model. Given the importance of retaining the balance between cross-sectional and temporal dimensions in the resample we implement the block resampling as follows. Without loss of generality, let $T = T(N)$ be some function of N. Then, the resample should be of temporal dimension $T(b)$ and cross-sectional dimension b, such that $b/T(b) \to c$. There exist $(T - T(b) + 1)(N - b + 1)$ distinct (but overlapping) blocks of consecutive observations of dimension $T(b) \times b$ in the data matrix Y. We can either resample all of them in the spirit of the block resampling scheme of Politis and Romano (1994), or if, as is usually the case, T and N are too large for resampling all the resamples to be computationally practical, we can resample a random selection of them without replacement.

As we have stated above we allow for an unknown sequence of normalizing constants, $\tau_{N,T}$. In fact, it is heuristically easy to see that, in the absence of a factor structure, the form of $\tau_{N,T}$ will depend on the upper tail of the limit of the ESD . To see this, we adapt a heuristic argument of Johnstone (2001). Let T be a function of N. Then, the constants $\tau_{N,T}$ are sole functions of N. Define $t_r = b^* - \hat{\mu}_r$ where b^* denotes the upper bound of the limit of the ESD. For any finite number, ℓ, let the ℓ smallest t_r lie in the interval $[0, g(N)]$. Denote the density associated with limiting distribution of the t_r by $f(x)$. Then, it is easy to see that

$$\int_0^{g(N)} f(x)\,dx \sim N^{-1}, \tag{5}$$

where $\sim$ denotes exact order behavior. The function $g(\cdot)$ that solves Equation (5) gives the order of magnitude of $\tau_{N,T}^{-1}$. It is clear then that $\tau_{N,T}$ are not easy to obtain analytically and depend crucially on f and g.

Subsampling can be used to estimate $\tau_{N,T}$. As discussed in chapter 8 of Politis, Romano, and Wolf (1999), if we define

$$L_{1,b,r}(x) = \frac{1}{N_b} \sum_{s=1}^{N_b} 1\{(\hat{\mu}^s_{r+1} - \hat{\mu}^s_{r^{\max}+1}) \le x\}, \qquad r = 0, \dots, r^{\max} - 1, \tag{6}$$

where the superscript ^{s} denotes that the sth subsample is used for the construction of the statistic, b denotes the subsample size, and $N_b = (T - T(b) + 1)(N - b + 1)$. Then, for any point $x > J_r(0)$

$$\log(L^{-1}_{1,b,r}(x)) = \log(J_r^{-1}(x)) - \log \tau_{b,T(b)} + o_p(1).$$

$L_{1,b,r}(x)$ can be viewed as a subsample estimator of the degenerate asymptotic distribution of $\hat{\mu}^s_{1+r} - \hat{\mu}^s_{r^{\max}+1}$. If we assume, as we do in Assumption 6, that $\tau_{N,T} = N^\beta$, $\beta > 0$ then, by estimating $J_r(x)$ using two different subsampling sizes $b_1 = N^{\beta_1}$ and $b_2 = N^{\beta_2}$, $1 > \beta_1 > \beta_2$, we get

$$\log\left(\frac{b_1}{b_2}\right)^{-1} \left(\log\left(L^{-1}_{1,b_1,r}(x)\right) - \log\left(L^{-1}_{1,b_2,j}(x)\right)\right) = \beta + o_p\left(\log\left(\frac{b_1}{b_2}\right)^{-1}\right). \tag{7}$$

Thus, an estimate of β can be obtained. To formalize this estimator we make the following assumption:

Assumption 6. $J_r(x)$ is continuous and strictly increasing in x. $\tau_{N,T}$ is of the form N^β $\beta > 0$.

The assumption that $\tau_{N,T(N)} = N^\beta$ can be dispensed with by assuming that $\tau_{N,T(N)} = h(N)$ for some increasing function $h(\cdot)$ with $\lim_{N\to\infty} h(N) = \infty$ by using remark 8.2.3 of Politis, Romano, and Wolf (1999). However, it is likely that estimating $h(\cdot)$ will be cumbersome in practice.

Let the number of subsampling replications be denoted by B, which, as discussed earlier, can be substantially smaller than the maximum possible number of subsampling replications, N_b. A final assumption on b and B which is needed for consistency of the estimator of the number of factors is the following:

Assumption 7. (i) $b \to \infty$, (ii) $b/N \to 0$, (iii) $B \to \infty$, and (iv) $bB/N \to 0$.

Of all the parts of Assumption 7 only part (iv) is nonstandard. This is needed, together with the last part of Assumption 4, to ensure that extreme subsampled statistics are smaller in order of magnitude than the original statistic. As mentioned in Remark 1, the last part of Assumption 4 is, in fact, quite mild, making it necessary for Assumption 7 to be relatively stringent. It is likely that a much weaker condition than that of Assumption 7(iv) will be sufficient for our theory to work. Nevertheless, we maintain this assumption so that we can relax Assumption 4, since both b and B are within our control. Denote by $\hat{L}_{T(b),b,r}(x)$ the subsampling estimate of the asymptotic distribution of $\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1})$. Then, we provide a formal definition of the new factor number estimator through the following algorithm:

Algorithm 1 (Estimation of number of factors).

Step 1. Demean the data $y_{i,t}$. Normalize $y_{i,t}$ by dividing every observation of each series with the estimated standard deviation of that series.

Step 2. Calculate the $r^{\max} + 1$ largest eigenvalues of the estimated covariance matrix of $y_{i,t}$, denoted $\hat{\mu}_r$, $r = 1, \dots, r^{\max}$.

Step 3. Set $r = 0$.

Step 4. Construct the test statistic $\hat{\mu}_{r+1} - \hat{\mu}_{r^{\max}+1}$. Using subsampling, estimate the normalizing constants $\tau^r_{N,T(N)}$ for this statistic. Denote these estimates by $\hat{\tau}^r_{N,T(N)}$.

Step 5. Compare $\hat{\tau}^r_{N,T(N)}(\hat{\mu}_{r+1} - \hat{\mu}_{r^{\max}+1})$ with the $1 - \alpha_N$ quantile of $\hat{L}_{T(b),b,r}(x)$, where $\alpha_N \to 0$.

Step 6. Set $\hat{r} = r$ if $\hat{\tau}^r_{N,T(N)}(\hat{\mu}_{r+1} - \hat{\mu}_{r^{\max}+1})$ does not exceed the quantile. Otherwise, set $r = r + 1$ and go to Step 4.

We refer to this algorithm as the MED (maximal eigenvalues distribution) algorithm. Note that the normalization carried out in Step 1 of the MED algorithm is a standard one used for most forms of factor analysis and does not have any theoretical implications. It simply implies that we focus our analysis on correlation matrices rather than covariance ones.

Remark 5. We briefly comment on the condition that $\alpha_N \to 0$ where $\alpha_N \equiv \alpha_{N,T(N)}$. Firstly we note that dependence on T is suppressed for notational convenience. The condition is imposed for consistent estimation of r^0 since otherwise with positive probability $\alpha > 0$, more factors than r^0 would be chosen by the sequential testing procedure. On the other hand, consistency of r^0 usually requires some condition on the rate at which $\alpha_N \to 0$ (see, e.g., proposition 3 of Hosoya 1989). In particular, that rate should not be too fast. Otherwise too few factors will be chosen. However, as our proofs show, the use of subsampling removes the need for such a condition. On the other hand, some alternative conditions are still needed, in this context, for consistency to be obtained. These conditions are given by Assumption 7(iv) and the last part of Assumption 4 which, as discussed in Remark 3, is, in our view, mild.

Then, we have the following theorems.

Theorem 1. Under Assumptions 1–5 and 7, and as $N, T \to \infty$, $\hat{L}_{T(b),b,r}(x)$ is a consistent estimator of the asymptotic distribution of $\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1})$, $r = r^0 + 1, \ldots, r^{\max}$.

Theorem 2. Under Assumptions 1–5 and 7, and as $N, T \to \infty$, $\hat{r}$ converges in probability to r^0.

We also have the following theorem on the estimated normalizing constants:

Theorem 3. Under Assumptions 1–7, for any $x > J_r(0)$, and setting

$$\hat{\beta} = \log\left(\frac{b_1}{b_2}\right)^{-1} \left(\log\left(L^{-1}_{1,b_1,r}(x)\right) - \log\left(L^{-1}_{1,b_2,r}(x)\right)\right),$$

$b_i = N^{\beta_i}$, $i = 1, 2$, $\hat{\tau}_{N,T} = N^{\hat{\beta}}$, we have that

$$\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}) - \hat{\tau}_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}) = o_p(1).$$

The proofs for these theorems are given in the Appendix.

Remark 6. Note that the need for a point x such that $x > J_r(0)$ to construct the above estimator is cumbersome rather than restrictive as a slightly more complicated argument can be used to dispense with this requirement as discussed in Politis, Romano, and Wolf (1999, ch. 8, pp. 181–182) and proven in Politis, Romano, and Wolf (1999, theorem 8.2.2).

Remark 7. The method suggested depends on two tuning parameters: b and α_N. A possible strategy allowed by the theory is to set the critical value for the test equal to the maximum of the B subsampling replications for the relevant test statistic. This implicitly fixes α_N to $1/B$. If $B \to \infty$ as $N \to \infty$ the theoretical condition is satisfied. A simpler alternative that we adopt in the Monte Carlo study and the empirical application is to let $\alpha_N = \alpha$. This implies that the number of factors is not consistently estimated, but is standard practice in analogous empirical circumstances. For example, when the cointegration rank of a multivariate time series is determined via standard eigenvalue based tests the significance level is usually kept fixed. On the choice of the block size b, there is little theory to guide this choice even in simpler setups. Politis, Romano, and Wolf (1999, ch. 9, sec. 9.4) discussed some possible solutions to this problem that might be useful in small samples in our context.

4. MONTE CARLO STUDY

4.1 Monte Carlo Setup

In this section we provide a detailed Monte Carlo study of the new number of factors estimator compared with the information criteria suggested by Bai and Ng (2002).

We now describe the Monte Carlo setup. The general model we consider has many similarities with Bai and Ng (2002) and is given by

$$\begin{aligned} y_{i,t} &= \sum_{j=1}^{r^0} \lambda_{j,i} f_{j,t} + \sum_{j=1}^{r^0} \lambda_{2,j,i} f_{j,t-1} + \epsilon_{i,t}, \\ &\qquad i = 1, \ldots, N, t = 1, \ldots, T, \\ \epsilon_t &= \Sigma^{1/2} \nu_t, \\ \nu_{i,t} &= \rho_i \nu_{i,t} + \xi_{i,t}. \end{aligned}$$

We set $f_{j,t} \sim \mathrm{N}(0,1)$, $\xi_{i,t} \sim \mathrm{N}(0, \theta r^0)$, $\lambda_{s,j,i} \sim \mathrm{N}(0,1)$, $N = 50, 100, 200$, $T = 50, 100, 200$.

One of the most important determinants of the performance of the number of factor estimators is the proportion of variance explained by the factors. This is controlled by θ. So, for example, in the case where $\rho_i = 0$ and $\Sigma = I$, $\theta = 1$ implies that R^2 is 0.66, whereas $\theta = 9$ implies that R^2 is 0.182. Evidence seems to suggest that in many datasets this R^2 is quite low. Hence, it is crucial that any method works well in these circumstances. We consider $\theta = 1, 9, 19$ leading to R^2 of 0.66, 0.182, and 0.095 when $\rho_i = 0$ and $\Sigma = I$. The latter value may seem extreme, but it will provide an envelope for the performance of the methods for most circumstances. Also we consider $r^0 = 2$ and $r^{\max} = 8$. For Experiments A, $\Sigma = I$, $\rho_i = 0$. For Experiments B, $\Sigma = [\sigma_{i,j}]$, $\sigma_{i,i} = 1$, $\sigma_{i,j} = \sigma_{j,i} \sim \mathrm{U}(-0.1, 0.1)$ for $|i - j| \leq 5$ and $\rho_i = 0.5$. Experiments C are as Experiments B, but $\rho_i = 0.95$. Finally, Experiments D are as Experiments C, but $\sigma_{i,j} = \sigma_{j,i} \sim \mathrm{U}(0, 0.199)$. This is the most extreme example of cross-sectional and temporal dependence we consider.

So the approximate factor models allow for considerable cross-sectional dependence and temporal dependence. For the MED algorithm we consider estimation of the convergence rate of the asymptotic distribution of the eigenvalues as discussed in Section 3. For subsampling we consider $b(N) = a(N)N$, where $a(N) = 0.7$ for $N = 50$, $a(N) = 0.6$ for $N = 100$, and $a(N) = 0.5$ for $N = 200$. Throughout this and the next section the significance level, α_N, is set to 0.01.

We compare the new method with the information criteria suggested by Bai and Ng (2002). These criteria, which are minimized over r, are given below:

$$PC_1(r) = V_r + r\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

$$PC_2(r) = V_r + r\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln C_{NT}^2,$$

$$PC_3(r) = V_r + r\hat{\sigma}^2\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right),$$

$$IC_1(r) = \ln(V_r) + r\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

$$IC_2(r) = \ln(V_r) + r\left(\frac{N+T}{NT}\right)\ln C_{NT}^2,$$

$$IC_3(r) = \ln(V_r) + r\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right),$$

where

$$V_r = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i-t} - \sum_{j=1}^{r}\hat{\lambda}_{j,i}\hat{f}_{j,t}\right)^2,$$

$C_{NT}^2 = \min(N, T)$ and $\hat{\sigma}^2 = V_{r^{\max}}$. Note that we choose to start the search at $r = 0$ both for the MED algorithm and for the information criteria. The Monte Carlo study of Bai and Ng (2002) did not consider the value $r = 0$ in the information criteria search. However, such a search does not address the very interesting problem of whether a given dataset supports a factor structure at all. Assuming the presence of at least one factor does not really seem as innocuous as usually presumed in the literature. Hence, we choose to modify the setting to address this very interesting question.

As a final point we note that, although the criteria developed by Hallin and Liska (2007) for generalized dynamic factor models are, strictly speaking, applicable to the simple dynamic factor data generation processes of our Monte Carlo study, we do not consider them in the Monte Carlo study. The main reason for this is that, as discussed in section 4 of Hallin and Liska (2007), the performance of these criteria can be greatly improved by a cross-validation procedure suggested by the authors related to the choice of constant in the penalty function. This cross-validation procedure has a number of steps for which a limited array of choices needs to be made by the researcher. A full and fair evaluation of these criteria would require an exploration of all these choices which, in our view, would be extensive and, as a result, detract from the main point of the current Monte Carlo study which is to evaluate our proposed method. However, since we see much of value in the suggested criteria of Hallin and Liska (2007) we consider them for our empirical applications discussed in Section 5.

Table 1. Experiments A

N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3
$\theta = 1$								
50	50	4.444	4.668	4.019	8.000	3.999	3.992	7.997
100	50	4.536	4.001	4.000	5.262	4.000	4.000	4.005
200	50	4.662	4.000	4.000	4.000	4.000	4.000	4.000
50	100	4.458	4.043	4.000	6.142	4.000	4.000	4.143
100	100	4.557	4.000	4.000	7.007	4.000	4.000	5.193
200	100	4.654	4.000	4.000	4.000	4.000	4.000	4.000
50	200	4.554	4.003	4.000	4.204	4.000	4.000	4.000
100	200	4.619	4.000	4.000	4.009	4.000	4.000	4.000
200	200	4.656	4.000	4.000	4.076	4.000	4.000	4.001
$\theta = 10$								
50	50	4.183	3.091	1.918	7.960	0.090	0.004	7.080
100	50	5.095	2.596	1.963	4.073	0.137	0.023	2.128
200	50	5.294	2.512	2.174	3.388	0.253	0.098	1.367
50	100	4.502	2.534	1.930	4.052	0.132	0.021	2.066
100	100	4.668	2.754	1.778	4.743	0.643	0.054	3.968
200	100	4.806	3.279	2.723	3.995	1.765	0.868	3.906
50	200	4.422	2.452	2.085	3.326	0.221	0.078	1.228
100	200	4.469	3.248	2.691	3.994	1.673	0.796	3.897
200	200	4.570	3.951	3.603	4.000	3.737	2.770	4.000
$\theta = 19$								
50	50	2.768	1.932	0.517	7.926	0.000	0.000	4.789
100	50	4.434	0.649	0.189	3.178	0.000	0.000	0.077
200	50	5.530	0.247	0.092	1.220	0.000	0.000	0.000
50	100	3.377	0.613	0.180	3.083	0.000	0.000	0.073
100	100	4.937	0.317	0.016	4.356	0.000	0.000	2.337
200	100	5.429	0.294	0.050	2.471	0.000	0.000	0.643
50	200	4.269	0.218	0.087	1.130	0.000	0.000	0.002
100	200	4.786	0.278	0.058	2.446	0.001	0.000	0.618
200	200	4.887	0.869	0.135	3.996	0.029	0.001	3.948

NOTE: MED refers to the method defined by Algorithm 1. PC_i and IC_i, $i = 1, 2, 3$, refer to the information criteria of Bai and Ng (2002).

4.2 Monte Carlo Results

Tables 1–4 report the average selected number of factors over 1000 replications for Experiments A–D. We start with results in Table 1. The setup here is one where the true number of static factors is equal to 4. For $\theta = 1$ all methods do well. In particular, the MED algorithm does particularly well for all experiments with the estimated number of factors, practically always being chosen to be slightly above 4. This is expected given that the test significance level is kept fixed. The information criteria do quite well too, with some problems being encountered at $N = 50$ for all values of T.

As soon as θ increases we note a marked deterioration in the performance of the information criteria. They underestimate the number of factors significantly in many cases. MED seems to suffer less and we conclude that it is relatively unaffected by the R^2 of the factor model unlike the other methods.

Moving on to Experiments B–D; in Tables 2–4 we see a steep deterioration of the performance of the information criteria. The pattern for Experiments B resembles that of Experiments A as θ rises. However, the introduction of cross-sectional and temporal dependence affects negatively the performance of information criteria. MED is again less affected. For Experiments C and D performance is similar. Information criteria now massively overestimate the number of factors. They practically always select the maximum allowable number of factors. Once again MED is the least affected providing reasonable estimates of the number of factors in all circumstances considered. In particular, while the estimated number of factors increases beyond the true number as N and T rise, the asymptotic results kick in, when at $N = 200$ a rise of T from 100 to 200 improves the estimate, both for Experiments C and D.

To conclude, MED seems to outperform the information criteria across a variety of Monte Carlo experiments. It seems insensitive to moderate cross-sectional and considerable temporal dependence. Importantly, it seems less sensitive to low R^2 for the factor equations compared to the information criteria. Given that factors are likely to explain a relatively small average proportion of the variance of empirical datasets due to the extreme parsimony of the factor model such a property is highly prized. The performance of MED makes the method a reasonable alternative to information criteria.

Table 2. Experiments B

N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3
$\theta = 1$								
50	50	4.890	7.129	5.758	8.000	4.785	3.986	8.000
100	50	4.966	6.762	5.825	8.000	4.404	4.043	8.000
200	50	5.188	6.632	5.990	7.967	4.211	4.036	7.634
50	100	4.589	5.423	4.591	7.945	4.012	4.000	7.552
100	100	4.741	4.617	4.016	8.000	4.000	4.000	8.000
200	100	4.785	4.176	4.007	7.906	4.000	4.000	7.508
50	200	4.601	4.128	4.033	5.305	4.000	4.000	4.020
100	200	4.631	4.000	4.000	5.947	4.000	4.000	4.278
200	200	4.814	4.000	4.000	8.000	4.000	4.000	8.000
$\theta = 10$								
50	50	3.058	5.776	3.942	8.000	0.128	0.002	8.000
100	50	4.899	5.302	4.117	7.989	0.176	0.026	7.744
200	50	5.983	5.174	4.423	7.382	0.211	0.086	2.147
50	100	3.912	3.434	2.433	7.046	0.033	0.001	2.498
100	100	5.447	3.025	1.735	8.000	0.198	0.005	8.000
200	100	6.193	3.109	2.428	7.248	0.698	0.235	5.185
50	200	4.548	2.151	1.742	3.405	0.019	0.005	0.406
100	200	5.365	2.640	2.010	4.490	0.461	0.103	3.439
200	200	5.580	3.609	2.703	8.000	2.332	0.806	7.998
$\theta = 19$								
50	50	2.275	5.449	3.567	8.000	0.040	0.000	8.000
100	50	3.559	4.865	3.599	7.936	0.033	0.002	7.068
200	50	4.811	4.658	3.917	6.821	0.016	0.002	0.889
50	100	2.158	2.654	1.454	6.484	0.000	0.000	0.557
100	100	3.350	1.618	0.282	8.000	0.000	0.000	8.000
200	100	4.798	1.200	0.455	6.400	0.000	0.000	1.715
50	200	2.621	0.418	0.164	1.945	0.000	0.000	0.001
100	200	4.517	0.212	0.027	3.281	0.000	0.000	0.316
200	200	5.915	0.424	0.036	7.995	0.002	0.000	7.968

NOTE: MED refers to the method defined by Algorithm 1. PC_i and IC_i, $i = 1, 2, 3$, refer to the information criteria of Bai and Ng (2002).

Table 3. Experiments C

N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3
$\theta = 1$								
50	50	3.617	8.000	8.000	8.000	8.000	7.999	8.000
100	50	6.101	8.000	8.000	8.000	8.000	8.000	8.000
200	50	7.656	8.000	8.000	8.000	8.000	8.000	8.000
50	100	4.053	8.000	8.000	8.000	8.000	8.000	8.000
100	100	6.470	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.457	8.000	8.000	8.000	8.000	8.000	8.000
50	200	3.559	8.000	8.000	8.000	8.000	8.000	8.000
100	200	5.874	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.296	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 10$								
50	50	1.760	7.999	7.852	8.000	7.990	7.306	8.000
100	50	4.039	8.000	7.999	8.000	8.000	7.969	8.000
200	50	6.863	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.733	8.000	7.997	8.000	8.000	7.989	8.000
100	100	6.211	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.769	8.000	8.000	8.000	8.000	8.000	8.000
50	200	4.070	8.000	8.000	8.000	8.000	8.000	8.000
100	200	6.392	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.727	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 19$								
50	50	1.563	8.000	7.852	8.000	7.993	7.401	8.000
100	50	4.020	7.999	7.995	8.000	7.999	7.970	8.000
200	50	6.773	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.459	8.000	8.000	8.000	8.000	7.998	8.000
100	100	6.017	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.726	8.000	8.000	8.000	8.000	8.000	8.000
50	200	4.041	8.000	8.000	8.000	8.000	8.000	8.000
100	200	6.234	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.744	8.000	8.000	8.000	8.000	8.000	8.000

NOTE: MED refers to the method defined by Algorithm 1. PC_i and IC_i, $i = 1, 2, 3$, refer to the information criteria of Bai and Ng (2002).

5. EMPIRICAL APPLICATIONS

We apply our suggested method of determining the number of factors to two large datasets. The first is a dataset of stock returns while the second is a dataset of macroeconomic variables. Before focusing on the empirical applications we note that we will consider the new method, the criteria of Bai and Ng (2002) and, in addition to methods investigated in the Monte Carlo study, the criteria of Hallin and Liska (2007). We, therefore, give a brief overview of these criteria including the implementation we consider for our empirical applications. Two criteria were proposed in Hallin and Liska (2007). They are given by

$$IC^T_{1:N}(r) = \frac{1}{N}\sum_{i=k+1}^{N} \frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T} \lambda^{*T}_{N,i}(\theta_l) + r\tilde{c}p(N,T), \qquad 0 \le r \le r^{\max},$$

and

$$IC^T_{2:N}(r) = \log\left[\frac{1}{N}\sum_{i=k+1}^{N} \frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T} \lambda^{*T}_{N,i}(\theta_l)\right] + r\tilde{c}p(N,T), \qquad 0 \le r \le r^{\max},$$

where $\lambda^{*T}_{N,i}(\theta)$ denotes the ith largest eigenvalue of the sample spectral density matrix $\Sigma^{*T}_N(\theta)$ given by

$$\Sigma^{*T}_N(\theta) = \frac{1}{2\pi}\sum_{u=-M_T}^{M_T} w(M_T^{-1}u)\Gamma^T_{Nu}e^{-iu\theta},$$

Γ^T_{Nu} is the sample cross-covariance matrix of y_t and y_{t-u}, $i = \sqrt{-1}$, $\theta_l = \pi l/(M_T + 1/2)$, M_T is a truncation parameter (set to $[0.5\sqrt{T}]$), $\tilde{c}$ is a positive constant, and $p(N,T)$ is a penalty function such that $p(N,T) \to 0$ and $Np(N,T) \to \infty$. Then, the number of factors chosen is equal to the value of r that minimizes the criteria. We denote this value by $\hat{r}^{HL,T,1}_N$ and $\hat{r}^{HL,T,2}_N$ for the two criteria, respectively. One suggestion for p that Hallin and Liska (2007) made is

$$p(N,T) = \left(M_T^{-2} + M_T^{1/2}T^{1/2} + N^{-1}\right) \times \log\left(\min\left[N, M_T^2, M_T^{-1/2}T^{1/2}\right]\right).$$

We use this since Hallin and Liska (2007) noted that the performance of the criteria does not crucially depend on this choice, or, for that matter, on the choice for M_T. What is crucial is the choice for $\tilde{c}$. Hallin and Liska (2007) suggested a cross-validation procedure for determining $\tilde{c}$ in section 4 of their article. Step (vi) of this cross-validation procedure involves two

Table 4. Experiments D

N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3
$\theta = 1$								
50	50	3.410	8.000	8.000	8.000	8.000	8.000	8.000
100	50	6.090	8.000	8.000	8.000	8.000	8.000	8.000
200	50	7.625	8.000	8.000	8.000	8.000	8.000	8.000
50	100	4.293	8.000	8.000	8.000	8.000	8.000	8.000
100	100	6.497	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.480	8.000	8.000	8.000	8.000	8.000	8.000
50	200	3.838	8.000	8.000	8.000	8.000	8.000	8.000
100	200	5.961	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.353	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 10$								
50	50	1.923	8.000	7.864	8.000	7.996	7.339	8.000
100	50	3.893	8.000	7.996	8.000	8.000	7.970	8.000
200	50	6.777	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.671	8.000	8.000	8.000	8.000	7.994	8.000
100	100	6.242	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.757	8.000	8.000	8.000	8.000	8.000	8.000
50	200	4.190	8.000	8.000	8.000	8.000	8.000	8.000
100	200	6.362	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.747	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 19$								
50	50	1.716	7.998	7.832	8.000	7.987	7.370	8.000
100	50	3.891	8.000	7.996	8.000	8.000	7.975	8.000
200	50	3.237	8.000	8.000	8.000	8.000	8.000	8.000
50	100	4.401	8.000	7.999	8.000	8.000	7.992	8.000
100	100	6.824	8.000	8.000	8.000	8.000	8.000	8.000
200	100	6.606	8.000	8.000	8.000	8.000	8.000	8.000
50	200	4.751	8.000	8.000	8.000	8.000	8.000	8.000
100	200	6.886	8.000	8.000	8.000	8.000	8.000	8.000
200	200	6.620	8.000	8.000	8.000	8.000	8.000	8.000

NOTE: MED refers to the method defined by Algorithm 1. PC_i and IC_i, $i = 1, 2, 3$, refer to the information criteria of Bai and Ng (2002).

Table 5. Eigenvalues of returns sample covariance matrix

Rank	Value	Rank	Value
1	72.408	11	1.883
2	9.465	12	1.800
3	5.960	13	1.720
4	5.722	14	1.604
5	5.003	15	1.603
6	4.157	16	1.519
7	2.744	17	1.470
8	2.548	18	1.436
9	2.298	19	1.398
10	2.060	20	1.379

choices. As the authors did not present a single choice as their preferred one we choose to focus on the graphical approach they suggested based on the curve $\tilde{c} \to S_{\tilde{c}}$, where

$$S_{\tilde{c}} = \left[(IJ)^{-1} \sum_{i,j} \left(\hat{r}_{N_j}^{HL,T_i} - (IJ)^{-1} \sum_{i,j} \hat{r}_{N_j}^{HL,T_i} \right)^2 \right]^{1/2},$$

where $0 < N_1 < \cdots < N_J \equiv N$ and $0 < T_1 < \cdots < T_I \equiv T$. Hallin and Liska (2007) suggested that the value of $\tilde{c}$ that corresponds to the second stability interval of the $S_{\tilde{c}}$ curve be chosen as the appropriate value of $\tilde{c}$.

For the first empirical application we choose the S&P500 dataset. The data, obtained from Datastream, are daily returns and span the period 01/01/1985 to 8/02/2005 comprising 5,245 observations. We choose to consider only companies for which data are available throughout the period leading us to have $N = 318$ for this dataset. Once all periods when markets were closed are dropped from the dataset the number of observations is 5,073.

We start with a preliminary analysis of the eigenvalues of the variance covariance matrix of the normalized returns data. The first 20 eigenvalues are reported in Table 5. The bound b suggested by Kapetanios (2004) is equal to 2.56 and if we were to use this bound and compare it to the eigenvalues in Table 5, we would choose seven factors. However, we know that this bound applies only to data without a factor structure and so we apply the formal methods discussed in previous sections. These are the MED method and the information criteria. We allow a maximum of 20 factors to be selected by any method. Table 6 presents the results of our analysis. We see from Column 1 of Table 6 that MED selects 15 factors which is a number considerably larger than that suggested by the information criteria methods which select at most six factors. In particular all information criteria methods select six factors apart from PC_5 which selects five factors. We next move on to consider the criteria proposed by Hallin and Liska (2007). Following the cross-validation approach summarized above, we graphically report the $S_{\tilde{c}}$ curve and the chosen number of factors as a function of $\tilde{c}$, in the first panel of Figure 1. For $S_{\tilde{c}}$ we set $N_1 = 118$, $N_j = N_1 + 50(j - 1)$, $j = 2, \ldots, 5$, $T_1 = 573$, $T_i = T_1 + 500$, $i = 2, \ldots, 10$ and report results for values of $\tilde{c}$ from 0.1 to 3 at intervals of 0.1. These choices correspond closely to similar choices made in Hallin and Liska (2007). $IC_{1:N}^{T}(r)$ chooses zero factors for all values of $c \geq 0.1$. We, therefore, focus on $IC_{2:N}^{T}(r)$ which, from the first panel of Figure 1, clearly indicates that the chosen number of factors is equal to one. We note that there appears to be little evidence of conditional mean dynamics in stock returns. This is evidenced, by, e.g., an average first order absolute autocorrelation of 0.034 across all stock returns. As a result, we choose to place more weight on the results of the other methods for determining the number of factors in this case.

We next look at the average R^2 over all stock returns when the number of factors selected by each method has been used to run the individual stock return regressions. In particular, we regress each stock return on the selected factors and take the average of the R^2 obtained, across all stock returns. These results are presented in the second column of Table 6. As we see the R^2

Table 6. Factor selection results

Method	No. of factors	R^2	Av. corr.
MED	15	0.361	0.020
PC_1	6	0.307	0.023
PC_2	6	0.307	0.023
PC_3	6	0.307	0.023
PC_4	6	0.307	0.023
PC_5	5	0.292	0.027
PC_6	6	0.307	0.023

NOTE: Av. corr. denotes average absolute correlation of the factor model residuals.

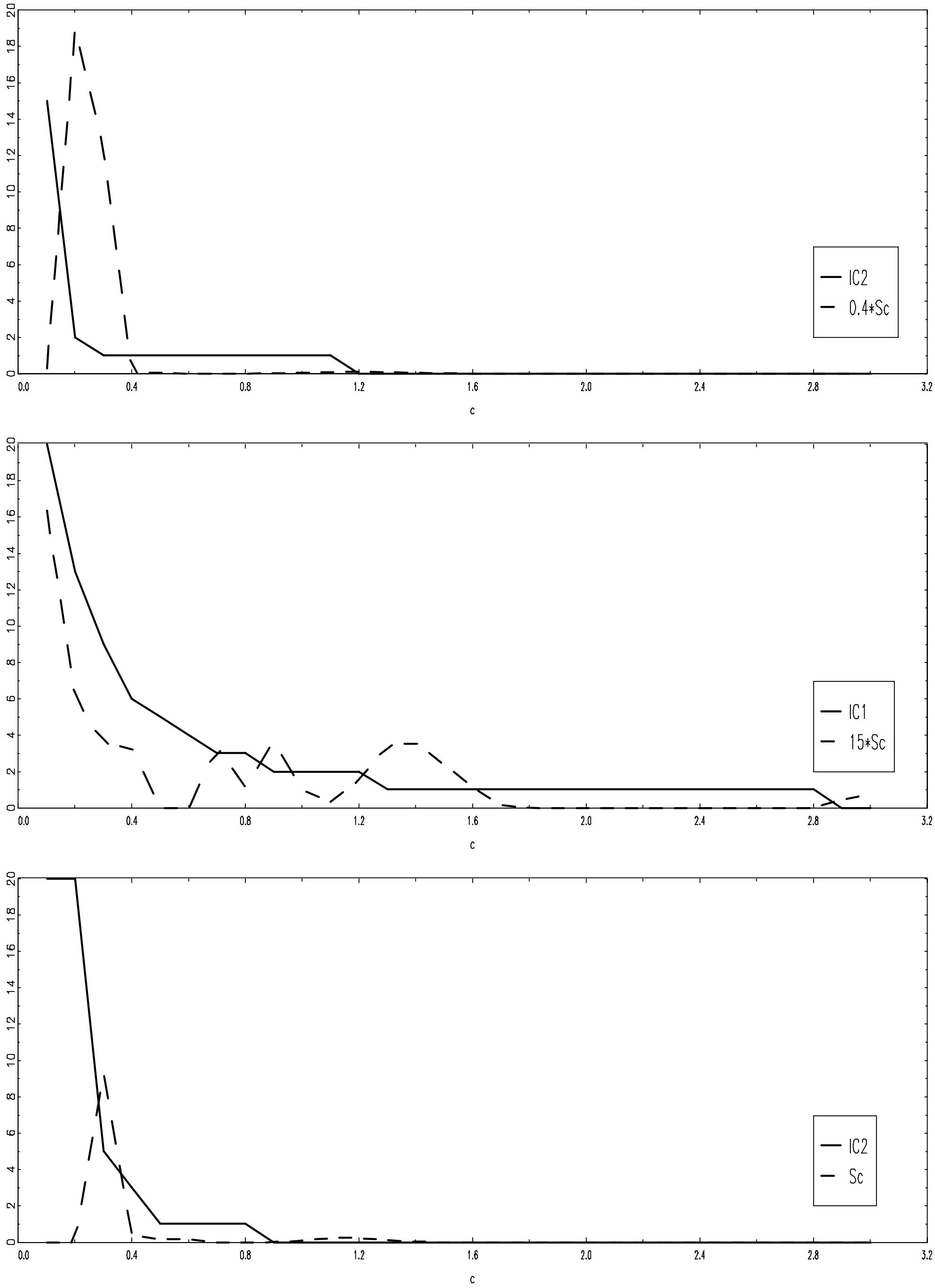

Figure 1. Graphical results for the Hallin and Liska (2007) information criteria. The first panel refers to the stock return data and $IC_{2:N}^{T}(r)$. The second and third panels refer to the macroeconomic variable data and $IC_{1:N}^{T}(r)$ and $IC_{2:N}^{T}(r)$, respectively.

obviously increases as we add more factors. If we look at the increase in R^2 as we successively add more factors there does not seem to be a discontinuity in the rate of increase. The presence of such a discontinuity has been traditionally used in the past to determine the number of factors. In fact, the only discontinuity appears for one factor which on its own explains about 22% of the variability in the returns data. Another notable feature is the fact that the R^2 is relatively low, compared to macroeconomic datasets where R^2 of 50%–60% are quite common even with only a couple of factors. This is the environment where information criteria are likely to underestimate the number of factors even for very large datasets with a large number of observations as the Monte Carlo study suggests. This seems to be reflected in our results.

A further issue concerns the properties of the idiosyncratic errors once the factors have been removed from the data. Although the factor model we entertain is likely to be an approximate factor model which allows for cross-sectional dependence of the idiosyncratic component, it is clearly reasonable to expect that most of the cross-sectional correlation is accounted for by the factors. To check this we calculate the average absolute cross-sectional correlation of the residuals once the appropriate number of factors according to each method has been removed from the data. The average correlation is reported in the third column of Table 6. One expects that when all the factors have been removed from the data then removing extra factors (principal components) will not reduce any further the residual cross-sectional correlation. However, as we see from Table 6, the average correlation is indeed reduced when 15 factors are removed, compared to the situation where only six factors are removed. To give an idea of the scale of the effect we note that the observed data have an average absolute correlation of 0.22 whereas once the first factor has been removed the average absolute correlation drops to 0.038.

For our second empirical application we focus on a large dataset of macroeconomic variables. This is taken from the work of Stock and Watson (2005). Their dataset consisted of monthly observations on 132 U.S. macroeconomic time series from 1959M1 through 2003M12. The predictors included series in 14 categories: real output and income, employment and hours, real retail, manufacturing and trade sales, consumption, housing starts and sales, real inventories, orders, stock prices, exchange rates, interest rates and spreads, money and credit quantity aggregates, price indexes, average hourly earnings, and miscellaneous. The series are transformed by taking logarithms and/or differencing so that the transformed series are approximately stationary. Specific transformations and the list of series is given in appendix A of Stock and Watson (2005).

We apply all methods for determining the number of factors in this dataset. Stock and Watson (2005) used the Bai and Ng (2002) criteria and reported that they chose seven factors. However, they did not make clear which criterion they chose. Examining their programs (available from *http://www.princeton.edu/~mwatson/ddisk/favar_ddisk.zip*) we find that they used IC_2. Setting the maximum number of factors to 20, we obtain the following results for PC_1, PC_2, PC_3, IC_1, IC_2, and IC_3, respectively: 17, 17, 20, 13, 7, and 20. Clearly, there is significant variation across the criteria. Our method suggests that there are 13 factors. We next move to the criteria of Hallin and Liska (2007). We focus on the graphical analysis of the second and third panels of Figure 1, which correspond to the two criteria, where we set $N_1 = 82$, $N_j = N_1 + 10(j-1)$, $j = 2, \dots, 6$, $T_1 = 227$, $T_i = T_1 + 20$, $i = 2, \dots, 16$ and report results for values of $\tilde{c}$ from 0.1 to 3 at intervals of 0.1. The first criterion clearly suggests one factor, whereas the second provides evidence for either one or zero factors. On balance the evidence from these criteria is for one factor. Unlike the stock return application, this macroeconomic dataset has important dynamics. As a result the conclusion of the Hallin and Liska (2007) criteria are more relevant. We conclude that there is evidence in favor of a generalized dynamic factor model with one factor.

The above two empirical applications make clear that the various methods are not necessarily substitutes but in many cases may complement each other. So, for the S&P 500 application, it is unlikely that the presence of conditional mean dynamics is relevant and so the methods of Hallin and Liska (2007) may be less useful than our methods. On the other hand for the macroeconomic dataset of Stock and Watson (2005), more focus should be placed on the results obtained using the methods of Hallin and Liska (2007).

6. CONCLUSIONS

Factor models for large datasets have gained much prominence in empirical and theoretical econometric work recently. Following on from the pathbreaking work of Stock and Watson (2002), a series of articles by Bai (2003, 2004) and Bai and Ng (2002) have provided the theoretical foundations of static factor models for large datasets. Work in Forni et al. (2000), and other articles by these authors, provided an alternative explicitly dynamic approach to factor analysis. An important issue in this work is choosing the number of factors to be included in the factor model. Other rigorous methods for doing this, in a variety of factor models, have been developed previously by Bai and Ng (2002, 2007), Amengual and Watson (2007), and Hallin and Liska (2007). All these methods are based on the use of information criteria.

This article suggests a new method for this problem. The method is based on the behavior of the eigenvalues of a large sample covariance matrix when no factor structure exists. In particular there exists a large literature on the fact that the largest eigenvalue of such a covariance matrix tends to a constant asymptotically. Since the behavior of the eigenvalues of the covariance matrix tend to infinity when a factor structure exists a method for distinguishing these two cases suggests itself. The article rigorously develops this idea, in the context of a sequential testing procedure for a variety of settings.

Monte Carlo analysis indicates that the method works well. In a majority of instances of empirical interest it outperforms some existing information criteria methods. Thus, it provides a useful alternative to existing methods.

APPENDIX: PROOFS OF THEOREMS

A.1 Proof of Theorem 1

Define

$$J_{T(N),N,r}(x,P) = \Pr_P\left\{\tau_{T(N),N}(\hat{\mu}_{r+1} - \hat{\mu}_{r^{\max}+1}) \le x\right\},$$
$$r = 0, \dots, r^{\max} - 1, \qquad \text{(A.1)}$$

where P denotes the unknown joint probability distribution of $\epsilon_{i,t}$, λ_i, and f_t. By Assumption 4, $J_{T(N),N,r}(x,P) \to J_r(x,P)$ as $N \to \infty$. The subsampling approximation to $J_{T(N),N,r}(x,P)$ is given by

$$L_{T(b),b,r}(x) = \frac{1}{N_b}\sum_{s=1}^{N_b} 1\left\{\tau_{T(b),b}(\hat{\mu}^s_{r+1} - \hat{\mu}^s_{r^{\max}+1}) \le x\right\},$$
$$r = 0, \dots, r^{\max} - 1. \qquad \text{(A.2)}$$

For x_α, where $J_r(x_\alpha, P) = \alpha$, we need to prove that $L_{T(b),b,r}(x_\alpha) \to J_r(x_\alpha, P)$ for the theorem to hold. But, by stationarity, $E(L_{T(b),b,r}(x_\alpha)) = J_{T(b),b,r}(x_\alpha, P)$ because as discussed in Section 3, the subsample is a sample from the true model, retaining both the temporal and cross-sectional ordering in the original sample. Hence, it suffices to show that $\mathrm{Var}(L_{T(b),b,r}(x_\alpha)) \to 0$ as $N \to \infty$. The subsampling approach we use samples blocks of data of dimension $T(b) \times b$ from the data matrix Y, and, therefore, neither the time series ordering nor the cross-sectional ordering is tampered with. Let

$$1_{b,s,r} = 1\{\tau_{T(b),b}(\hat{\mu}^s_{r+1} - \hat{\mu}^s_{r^{\max}+1}) \le x_\alpha\}, \qquad \text{(A.3)}$$

$$v_{N_b,h,r} = \frac{1}{N_b}\sum_{s=1}^{N_b} \mathrm{Cov}(1_{b,s,r}, 1_{b,s+h,r}). \qquad \text{(A.4)}$$

Then,

$$\begin{aligned}\mathrm{Var}(L_{T(b),b,r}(x_\alpha)) &\\ &= \frac{1}{N_b}\left(v_{N_b,0,r} + 2\sum_{h=1}^{N_b-1} v_{N_b,h,r}\right)\\ &= \frac{1}{N_b}\left(v_{N_b,0,r} + 2\sum_{h=1}^{b-1} v_{N_b,h,r}\right) + \frac{2}{N_b}\sum_{h=b}^{N_b-1} v_{N_b,h,r}\\ &= V_1 + V_2. \qquad \text{(A.5)}\end{aligned}$$

We first determine the order of magnitude of V_1. By the boundedness of $1_{b,s,r}$, it follows that $v_{N_b,h,r}$ is uniformly bounded across h. Hence, $|V_1| \le \frac{b}{N_b}\max_h |v_{N_b,h,r}|$, from which it follows that $V_1 = O(b/N_b) = o(1)$. We next examine V_2. For this we need to note that

$$|V_2| \le \frac{2}{N_b}\sum_{h=b}^{N_b-1} |v_{N_b,h,r}|, \qquad \text{(A.6)}$$

But, by the mixing assumption of Assumption 2 and Lemma 1, it then follows that $v_{N_b,h,r} = o(1)$ as $h \to \infty$. Hence,

$$V_2 = \frac{2}{N_b}\sum_{h=b}^{N_b-1} |v_{N_b,h,r}| = o(1),$$

proving the convergence of $L_{T(b),b,r}(x_\alpha)$ to $J_r(x_\alpha, P)$.

In order to complete the proof we finally need to show that

$$\hat{L}_{T(b),b,r}(x_\alpha) = \frac{1}{B}\sum_{s=1}^{B} 1\{\tau_{T(b),b}(\hat{\mu}^s_{r+1} - \hat{\mu}^s_{r^{\max}+1}) \le x\}, \qquad j = 0, \ldots, r^{\max},$$

converges in probability to $L_{T(b),b,r}(x_\alpha)$ as $B \to \infty$. But this result follows from proposition 4.1 of Romano (1989).

A.2 Proof of Theorem 2

In order to prove the theorem we need to show firstly that if $H_{0,r}: r^0 = r$, $r = 0, \ldots, r^{\max} - 1$ holds then

$$\Pr_P(\tau_{T(N),N}(\hat{\mu}_{r+1} - \hat{\mu}_{r^{\max}+1}) > \hat{q}^b_{r,\alpha,N}) = o(1), \qquad \text{(A.7)}$$

where $\hat{q}^b_{r,\alpha_N}$ solves

$$1 - L_{T(b),b,r}(x) = \alpha_N, \qquad \text{(A.8)}$$

and secondly that, if $H_{1,r}: r^0 > r$ holds then

$$\lim_{N\to\infty} \Pr_P(\tau_{T(N),N}(\hat{\mu}_{r+1} - \hat{\mu}_{r^{\max}+1}) > \hat{q}^b_{r,\alpha,N}) = 1. \qquad \text{(A.9)}$$

In the case where Equation (A.8) has a continuum of solutions we choose the minimum value of x such that Equation (A.8) holds. But Equation (A.7) holds if

$$\lim_{N\to\infty} \Pr_P(\tau_{T(N),N}(\hat{\mu}_{r+1} - \hat{\mu}_{r^{\max}+1}) \le \hat{q}^b_{r,\alpha,N}) = J_r(\hat{q}^b_{r,\alpha,N}, P) = 1, \qquad \text{(A.10)}$$

By Theorem 1 we have that $L_{T(b),b,r}(x_\alpha) \to J_r(x_\alpha, P)$ for all x. Thus, Equation (A.10), and, therefore, Equation (A.7), followed by Equation (A.8), and $\alpha_N \to 0$. We now have to show that Equation (A.9) holds. We establish the following three facts. Firstly, by Weyl's Theorem [see, e.g., Lutkepohl 1996, 5.3.2(9)], $\hat{\mu}_{r+1} \ge \bar{\mu}_{r+1}$, $r = 0, \ldots, r^{\max} - 1$ where $\bar{\mu}_{r+1}$ is the $r+1$ largest eigenvalue of $\Lambda' F' F \Lambda$ which is $O_p(N)$. Secondly, again by Weyl's theorem it follows that $\hat{\mu}_{r^{\max}+1}$ is smaller than the largest eigenvalue of $1/T\epsilon'\epsilon$ and is, therefore, bounded in probability. Thirdly, note that $\hat{q}^b_{\alpha_N}$ is at most equal to the maximum of all subsampled statistics. But, this maximum is, at most, $O_p(B^\kappa \tau_{T(b),b} b)$ by the last part of Assumption 4. But, $O_p(B^\kappa \tau_{T(b),b} b) = o_p(N\tau_{T(b),b}) = o_p(N\tau_{T(N),N})$ where the first equality follows from part (iv) of Assumption 7. The first two facts imply that $\tau_{T(N),N}(\hat{\mu}_{1+r} - \hat{\mu}_{r^{\max}+1})$ tends to infinity under $H_{1,r}$ at rate $\tau_{T(N),N} N$. This together with the third fact imply Equation (A.9).

A.3 Proof of Theorem 3

It is sufficient to show the following: Firstly

$$\tau_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}) - \hat{\tau}_{N,T}(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}) = o_p(1), \qquad \text{(A.11)}$$

if $\hat{\beta} - \beta = o_p((\ln N)^{-1})$ and secondly that

$$\hat{\beta} - \beta = o_p((\ln N)^{-1}). \qquad \text{(A.12)}$$

But, by Equation (7), we get that $\hat{\beta} - \beta = o_p(\log(\frac{b_1}{b_2})^{-1})$ and since $\log(\frac{b_1}{b_2})^{-1} = o((\ln N)^{-1})$ we get Equation (A.12). To get Equation (A.11) we simply note that it is sufficient to show that

$$\hat{\tau}_{N,T}/\tau_{N,T} = N^{\hat{\beta}}/N^{\beta} = 1 + o_p(1), \qquad \text{(A.13)}$$

since $(\hat{\tau}_{N,T} - \tau_{N,T})(\hat{\mu}_r - \hat{\mu}_{r^{\max}+1}) = o_p(1)$ if $\hat{\tau}_{N,T} - \tau_{N,T} = o_p(\tau_{N,T})$ or $\hat{\tau}_{N,T}/\tau_{N,T} = 1 + o_p(1)$. But $N^{\hat{\beta}}/N^{\beta} = 1 + o_p(1)$ if $\ln(N^{\hat{\beta}}) - \ln(N^{\beta}) = o_p(1)$ or $(\hat{\beta} - \beta)\ln N = o_p(1)$.

ACKNOWLEDGMENTS

The authors thank the current Co-Editor, Serena Ng, the retired Editor, Torben Andersen, an Associate Editor, and four anonymous Referees for their extremely helpful and insightful comments that greatly improved the article.

[Received September 2007. Revised November 2007.]

REFERENCES

Amengual, D., and Watson, M. W. (2007), "Consistent Estimation of the Number of Dynamic Factors in a Large *N* and *T* Panel," *Journal of Business & Economic Statistics*, 25 (1), 91–96. [397,399,407]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dinensions," *Econometrica*, 71, 135–173. [407]

——— (2004), "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics*, 122 (1), 137–183. [407]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [397,399,402-405,407]

——— (2007), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, 25 (1), 52–60. [397,399, 407]

Bai, Z. D., and Silverstein, J. W. (1998), "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Random Matrices," *The Annals of Probability*, 26 (1), 316–345. [398,400]

Baik, J., Ben-Arous, G., and Peche, S. (2004), "Phase Transition of the Largest Eigenvalue for Non-Null Complex Sample Covariance Matrices," mimeo, available at *http://arxiv.org/abs/math.PR/0403022*. [398]

Camba-Mendez, G., and Kapetanios, G. (2005), "Estimating the Rank of the Spectral Density Matrix," *Journal of Time Series Analysis*, 26 (1), 37–48. [399]

Camba-Mendez, G., Kapetanios, G., Smith, R. J., and Weale, M. R. (2003), "Tests of Rank in Reduced Rank Regression Models," *Journal of Business & Economic Statistics*, 21 (1), 145–155. [399]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324. [398]

Connor, G., and Korajczyk, R. A. (1993), "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–1292. [400]

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press. [400]

Doukhan, P. (1994), *Mixing: properties and examples*, Berlin: Springer-Verlag. [400]

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalised Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554. [397,399,407]

——— (2004), "The Generalized Dynamic Factor Model: Consistency and Rates," *Journal of Econometrics*, 119 (2), 231–255. [397,399]

Galambos, J. (1978), *The Asymptotic Theory of Extreme Order Statistics*, Wiley. [400]

Hachem, W., Loubaton, P., and Najim, J. (2005a), "The Empirical Eigenvalue Distribution of a Gram Matrix: From Independence to Stationarity," mimeo, Ecole Superieur d'Electricite, available at *http://arxiv.org/abs/math.PR/0502535*. [398]

——— (2005b), "The Empirical Eigenvalue Distribution of a Gram Matrix With a Given Variance Profile," mimeo, Ecole Superieur d'Electricite, available at *http://arxiv.org/abs/math.PR/0411333*. [398]

Hallin, M., and Liska, R. (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102, 603–617. [397,399,403-407]

Hosoya, Y. (1989), "Hierarchical Statistical Models and a Generalised Likelihood Ratio Test," *Journal of the Royal Statistical Society, Ser. B*, 51 (3), 435–448. [402]

Jacobs, J. P. A. M., and Otter, P. W. (2006), "Determining the Number of Factors and Lag Order in Dynamic Factor Models: A Minimum Entropy Approach," *Econometric Reviews*, to appear. [397]

Johnstone, I. M. (2001), "On the Distribution of the Largest Eigenvalue in Principal Component Analysis," *The Annals of Statistics*, 29, 295–327. [398,401]

Kapetanios, G. (2004), "A New Method for Determining the Number of Factors in Factor Models With Large Datasets," Working Paper 525, Queen Mary, University of London. [397,405]

Lewbel, A. (1991), "The Rank of Demand Systems: Theory and Nonparametric Estimation," *Econometrica*, 59, 711–730. [397]

Lutkepohl, H. (1996), *Handbook of Matrices*, New York: Wiley. [408]

Mahalanobis, P. (1946), "Sample Surveys of Crop Yields in India," *Sankya, Ser. A*, 7, 269–280. [400]

Onatski, A. (2005), "Determining the Number of Factors From Empirical Distribution of Eigenvalues," mimeo, Columbia University. [397]

——— (2006), "A Formal Statistical Test for the Number of Factors in the Approximate Factor Models," mimeo, Columbia University. [399,400]

Peche, S. (2005), "Universality of Local Eigenvalue Statistics for Random Sample Covariance Matrices," Ph.D. thesis, Ecole Polytechnique de Lausanne. [398]

Politis, D. N., and Romano, J. P. (1994), "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *The Annals of Statistics*, 22, 2031–2050. [400,401]

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*, Berlin: Springer-Verlag. [400-402]

Romano, J. P. (1989), "Bootstrap and Randomization Tests of Some Nonparametric Hypotheses," *The Annals of Statistics*, 17, 141–159. [408]

Stock, J. H., and Watson, M. W. (1989), "New Indices of Coincident and Leading Indicators," in *NBER Macroeconomics Annual 1989*, eds. O. J. Blanchard and S. Fischer, Cambridge: MIT Press. [397]

——— (2002), "Macroeconomic Forecasting Using Diffusion Indices," *Journal of Business & Economic Statistics*, 20, 147–162. [397,407]

——— (2005), "Implications of Dynamic Factor Models for VAR Analysis," mimeo, Princeton University. [407]

Tracy, C. H., and Widom, H. (1996), "On Orthogonal and Symplectic Matrix Ensembles," *Communications in Mathematical Physics*, 177, 727–754. [398]

Yin, Y. Q., Bai, Z. D., and Krishnaiah, P. R. (1988), "On the Limit of the Largest Eigenvalue of the Large Dimensional Sample Covariance Matrix," *Probability Theory and Related Fields*, 78, 509–521. [398]