

Department of Economics

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

George Kapetanios

Working Paper No. 551

December 2005

ISSN 1473-0278



Queen Mary
University of London

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

George Kapetanios*
Queen Mary, University of London.

December 16, 2005

Abstract

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 modelled 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 modelling. This paper provides builds on the work of Kapetanios (2004) to provide 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 favourable for the new method and suggest that it is a reasonable alternative to existing methods.

Keywords: Factor Models, Large Sample Covariance Matrix, Maximum Eigenvalue
JEL Codes: C12, C15, C23

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 modelled 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 been focused on small datasets meaning that the number of variables, N , to be modelled via a factor model is finite. Recently, Stock and Watson

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS. email: G.Kapetanios@qmul.ac.uk.

(2002) have put forward the case for analysing large datasets via factor analysis, where N is allowed to tend to infinity. Stock and Watson (2002) suggest the use of principal components for estimating factors in this context. Further work has been carried out by, e.g., Forni, Hallin, Lippi, and Reichlin (2000) and Forni, Hallin, Lippi, and Reichlin (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 modelling. The only 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 .

This paper 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 much 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 the author (Kapetanios (2004)) has made use of random matrix theory (RMT) to devise methods for determining the number of factors in large datasets. However, a number of problems existed with the approach suggested in that paper. The main problem related to the stringency of the assumptions made to derive formal results for the method. In the current paper 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 i.i.d. 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 paper is organised as follows: Section 2 surveys the available results on the behaviour of the eigenvalues of large 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. Finally, Section 5 concludes.

2 Preliminaries

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

$$y_{i,t} = f_t' \lambda_i + \epsilon_{i,t} \quad (1)$$

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 with mild cross-sectional dependence. The nature of this dependence will be made clear later.

Rewriting the above model in matrix notation gives

$$Y = F\Lambda + \epsilon \quad (2)$$

where $Y = (Y_1, \dots, Y_N)$, $F = (F_1, \dots, F_r)$, $\Lambda = (\lambda_1, \dots, \lambda_N)$, $\epsilon = (\epsilon_1, \dots, \epsilon_N)$, $Y_i = (y_{i,1}, \dots, y_{i,T})'$, $F_i = (f_{i,1}, \dots, f_{i,T})'$ 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 variance covariance matrix of the dataset is given by

$$\Sigma_Y = \Sigma_f + \Sigma_\epsilon \quad (3)$$

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 i.i.d. 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 $c = \lim_{N,T \rightarrow \infty} \frac{N}{T}$. The result is remarkable in its simplicity. For example, for $N = T$ the largest eigenvalue converges almost surely to 4. This result has been proven repeatedly 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.

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 $N, T \rightarrow \infty$ 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 latter support, of course, depends on c .

The above results relate to temporally i.i.d. data. Recently, work by Hachem, Loubaton, and Najim (2005a) and Hachem, Loubaton, and Najim (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 MA coefficients 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 rarer. 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 a 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

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

where $\hat{\mu}_1$ is the maximum eigenvalue of the sample covariance matrix of an $T \times N$ matrix of i.i.d. $N(0, 1)$ variates. There does not seem to be any work publicly available on convergence rates for temporally or cross-sectionally dependent data.

We now outline the suggested estimation method for the number of factors. It clear that if the number of factors in the dataset is r then the first r eigenvalues of Σ_Y will increase at rate N whereas the rest will remain bounded. It is reasonable to expect a similar behaviour 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 by $\hat{\mu}_i$, $i = 1, \dots, N$. Then it is reasonable to expect that for some $r < rmax$, $\hat{\mu}_i - \hat{\mu}_{rmax+1}$ will tend to infinity for $i = 1, \dots, r^0$ but remain bounded for $i = r^0 + 1, \dots, rmax$. The role of $\hat{\mu}_{rmax+1}$ is as an estimator of the upper bound for the maximum eigenvalue when there is no factor structure in the dataset. For example, for i.i.d. 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 spelt out in the next section, $O(1)$ of the largest eigenvalues will tend to the upper bound of the ESD and so $\hat{\mu}_{rmax+1}$ is a valid estimator for that bound.

If there is no factor structure then $\hat{\mu}_i - \hat{\mu}_{rmax+1}$, $i = 1, \dots, rmax$, suitably normalised 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 $\tau_{N,T}$ were 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 $\hat{\mu}_{r+1} - \hat{\mu}_{rmax+1}$. 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 $\hat{\mu}_{r+1} - \hat{\mu}_{rmax+1}$ can be approximated and a test can be carried out. Then a sequence of such tests can be used to determine the 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 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, Kapetanios, Smith, and Weale (2003) and Camba-Mendez and Kapetanios (2005).

3 Theory

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

Assumption 1 $T^{-1} \sum_{t=1}^T f_t f_t' \xrightarrow{p} \Sigma$ for some $r \times r$ positive definite matrix Σ .

Assumption 2 $E(\lambda_i \lambda_i) = D$ for some positive definite matrix D . There is an ordering of the cross-sectional units 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 $\zeta > 0$

Assumption 3 $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = \sigma_i^2$, $E|\epsilon_{i,t}|^4 \leq M$. The largest eigenvalue of $E(1/T \epsilon' \epsilon)$ is bounded.

Assumption 4 Let r^0 denote the number of true factors. Then, for every $i = r^0 + 1, \dots, rmax$, there exists a sequence of constants $\tau_{N,T}$ such that $\tau_{N,T}(\hat{\mu}_i - \hat{\mu}_{rmax+1}) \xrightarrow{d} J_i$ where J_i is a limit law.

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

These assumptions are less restrictive than those used in Bai and Ng (2002). Assumptions 1 and 3 impose minimal conditions on the factors and idiosyncratic errors. Assumption 4 is not standard but simply posits the existence of a limit law for the normalised

difference of the eigenvalues. No assumptions are placed on that limit law such as continuity. Further nothing is assumed about the rate of convergence to that limit law or even whether $\tau_{N,T}$ actually tends to infinity. Notice that no conditions are explicitly placed on the temporal or cross-sectional dependence of $\epsilon_{i,t}$ either. Of course, some conditions are implied by the need for bounded eigenvalues for Σ_ϵ . 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 $\hat{\Sigma}_\epsilon$. 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 $\delta > 1/(\zeta - 1)$ where $E(|\epsilon_{i,t}|^{2+2\delta}) < \infty$.*

The lemma is proven in the appendix. Note that the above Lemma relaxes Theorem 1 of Connor and Korajczyk (1993) which requires that $\delta > 2/(\zeta - 2)$. In this context, it is worth noting that the maximum eigenvalue of the sample covariance matrix of the estimator of the idiosyncratic component $\epsilon_{i,t}$ obtained via the method of principal components is bounded. The following Lemma formalises this assertion and is proven in the appendix.

Lemma 2 *Let $\hat{\epsilon}$ denote the estimator of ϵ using principal components. Then, under assumptions 1-5 and assumptions A-D of Bai (2003) the maximum eigenvalue of $1/T\hat{\epsilon}'\epsilon$ is $O_p(1)$.*

We next discuss the subsampling methodology for estimating J_i . 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 the resampled samples are smaller 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 data are both cross-sectionally and temporally dependent. In these cases block resampling is suggested by Politis and Romano (1994). However, in our case there is an asymmetry between temporal and cross-sectional resampling. The temporal ordering is clearly of importance and needs to be retained when resampling. On the other hand the cross-sectional ordering is unknown in the case of the ϵ_i and irrelevant in the case of the factor structure since this structure is retained intact when units are reordered. We therefore resample whole individual units (y_i) without replacement to carry out subsampling. More specifically letting $T(N)$ be a function of N , we resample $b \ll N$ units without replacement and for each unit we retain only observations indexed $t, \dots, t+T(b)$ for some random observation $t \leq T(N) - T(b)$.

As we have stated above we allow for an unknown sequence of normalising 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 $t_i = b - \hat{\mu}_i$ where b denotes the upper bound of the limit of the ESD. Let the $O(1)$ smallest t_i lie in the interval $[0, g(N)]$. Denote the density associated with limiting distribution of the t_i by $f(x)$. Then, it is easy to see that

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

where \sim denotes exact order behaviour. The function $g(\cdot)$ that solves (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

$$L_{0,b,i}(x) = \frac{1}{N_b} \sum_{j=1}^{N_b} 1 \{ (\hat{\mu}_{1+i}^j - \hat{\mu}_{r^{max}+1}^j) \leq x \}, \quad i = 0, \dots, r^{max} - 1 \quad (6)$$

where the superscript j denotes the j -th subsample and b denotes the subsample size, then for any point $x > J_i(0)$

$$\log(L_{0,b,i}^{-1}(x)) = \log(J_i^{-1}(x)) - \log \tau_{b,T(b)} + o_p(1)$$

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

$$\log \left(\frac{b_1}{b_2} \right)^{-1} (\log(L_{0,b_1,i}^{-1}(x)) - \log(L_{0,b_2,i}^{-1}(x))) = \beta + o_p \left(\log \left(\frac{b_1}{b_2} \right)^{-1} \right) \quad (7)$$

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

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

Note that the need for a point x such that $x > J_i(0)$ 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 p.

181). Further, 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 $\lim_{N \rightarrow \infty} h(N) = \infty$ 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.

Denote by $L_{T(b),b,i}(x)$ the subsampling estimate of the asymptotic distribution of $\tau_{N,T}(\hat{\mu}_i - \hat{\mu}_{rmax+1})$. 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 $y_{i,t}$. Normalise $y_{i,t}$ by dividing every observation of each series with the estimated standard deviation of that series.

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

Step 3 Set $i = 0$.

Step 4 Construct the test statistic $\hat{\mu}_{i+1} - \hat{\mu}_{rmax+1}$. Using subsampling, estimate the normalising constants $\tau_{N,T(N)}^i$ for this statistic.

Step 5 Compare $\hat{\tau}_{N,T(N)}^i(\hat{\mu}_{i+1} - \hat{\mu}_{rmax+1})$ with the $1 - \alpha_N$ quantile of the subsampling distribution given by $L_{T(b),b,i}(x)$ where $\alpha_N \rightarrow 0$.

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

We refer to this algorithm as the MED (maximal eigenvalues distribution) algorithm. Then, we have the following theorems.

Theorem 1 *Under assumptions 1-5, $b = o(N)$, and as $N, T \rightarrow \infty$, $L_{T(b),b,i}(x)$ is a consistent estimator of the asymptotic distribution of $\tau_{N,T}(\hat{\mu}_i - \hat{\mu}_{rmax+1})$, $i = r^0 + 1, \dots, rmax$*

Theorem 2 *Under assumptions 1-5, and as $N, T \rightarrow \infty$, \hat{r} converges in probability to r^0 .*

We also have the following theorem on the estimated normalising constants.

Theorem 3 *Under assumptions 1-6 and setting*

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

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

$$\tau_{N,T}(\hat{\mu}_i - \hat{\mu}_{rmax+1}) - \hat{\tau}_{N,T}(\hat{\mu}_i - \hat{\mu}_{rmax+1}) = o_p(1)$$

The proofs for these theorems are given in the Appendix.

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 also consider the method of Kapetanios (2004) referred to as the ME (maximum eigenvalue) algorithm.

The general model we consider has many similarities with Bai and Ng (2002) and is given by

$$y_{i,t} = \sum_{j=1}^{r^0} \lambda_{j,i} f_{j,t} + \sum_{j=1}^{r^0} \lambda_{2,j,i} f_{j,t-1} + \epsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

$$\epsilon_t = \Sigma^{1/2} \nu_t$$

$$\nu_{i,t} = \rho_i \nu_{i,t} + \xi_{i,t}$$

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

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 $\theta = 1$, R^2 is 0.66 whereas for $\theta = 9$, 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, 0.1 and 0.095. 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 $rmax = 8$. For Experiments A, $\Sigma = I$, $\rho_i = 0$. For experiments B, $\Sigma = [\sigma_{i,j}]$, $\sigma_{i,i} = 1$, $\sigma_{i,j} = \sigma_{j,i} \sim U(-0.1, 0.1)$ 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 $\sigma_{i,j} = \sigma_{j,i} \sim U(0, 0.199)$. This is the most extreme example of cross-sectional and temporal dependence we consider. We should note two points. Firstly, for the approximate factor models the values of R^2 reported above are not exactly accurate

because of the AR structure of $\nu_{i,t}$ which implies lower values for R^2 than reported. Secondly, the calculation giving rise to the reported values of R^2 take into account both the variability of $f_{j,t}$ and $\lambda_{j,i}$ which are assumed random.

So the approximate factor models allow for considerable cross-sectional dependence and temporal dependence. For the MED algorithm we do not consider estimation of the convergence rate of the asymptotic distribution of the eigenvalues but simply normalise by the correct rate for i.i.d. data given in (4). For subsampling we consider $b(N) = a(N)N$ where $a(N) = 0.7, 0.6, 0.5$ for $N = 50, 100, 200$. Finally, each test of the sequence of tests for MED is carried out at the 1% significance level. These choices are made for simplicity and to illustrate that the approach can still work when used in a simplified setting. We compare the new method with the information criteria suggested by Bai and Ng (2002). These criteria, which are minimised over r , are given below

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

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

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

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

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

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

where

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

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.

4.2 Monte Carlo Results

Tables 1-4 report the average selected number of factors over 1000 replications for Experiments A-D. Results make interesting reading. 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 . The method suggested in Kapetanios (2004) also performs very well.

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. ME seems to suffer as well but to a much lesser extent. MED seems to suffer least 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. ME is affected as well overestimating the number of factors. MED is again least affected. For experiments C and D performance is similar. Information criteria now overestimate massively the number of factors. They select the maximum allowable number of factors practically always. A similar result appears for ME which is expected as this method cannot deal with temporal dependence. 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 for $\theta = 1$ 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. Nevertheless, we still feel that the ease of implementation of ME together with the favourable Monte Carlo results reported in Kapetanios (2004) allow it to be a reasonable competitor to MED and information criteria.

5 Conclusions

Factor models for large datasets have gained much prominence in empirical and theoretical econometric work recently. Following on from the path breaking work of Stock and Watson (2002) a series of papers by Bai and Bai and Ng (Bai and Ng (2002), Bai (2003), Bai (2004)) have provided the theoretical foundations of static factor models for large datasets. Work in Forni, Hallin, Lippi, and Reichlin (2000) and other papers by these authors have 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. The only rigorous method for doing this has been developed in an influential paper by Bai and Ng (2002) and uses information criteria.

This paper suggests a new method for this problem. The method is based on the behaviour 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 tend to a constant asymptotically. Since the behaviour 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 paper develops rigorously this idea for a variety of settings following on the work of Kapetanios (2004).

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

Appendix

5.1 Proof of Lemma 1

We need to derive a trade-off between the order of moments that need to exist and the α -mixing size for the sequence $\{\epsilon_{i,t}\}_{i=1}^N$ for all t such that the covariance matrix of $\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{N,t})'$ has bounded eigenvalues. By Schwarz, Rutishauser, and Stiefel (1973), the eigenvalues of Σ_ϵ will be bounded if the column sum norm of Σ_ϵ is bounded. This will be the case if the covariances $E(\epsilon_{i,t}\epsilon_{i+m,t})$ of ϵ_t are absolutely summable. This will be the case if $|E(\epsilon_{i,t}\epsilon_{i+m,t})| = O(m^{-s})$ as $m \rightarrow \infty$ where $s > 1$. By Corollary 14.3 of Davidson (1994)

$$|E(\epsilon_{i,t}\epsilon_{i+m,t})| \leq C\alpha_m^{1-2/(2+2\delta)}$$

where $E(|\epsilon_{i,t}|^{2+2\delta}) < \infty$. Thus $|E(\epsilon_{i,t}\epsilon_{i+m,t})| = O(m^{-\zeta(1-1/(1+\delta))})$. So we need

$$\zeta(1 - 1/(1 + \delta)) > 1$$

or $\delta > 1/(\zeta - 1)$. Clearly, this is a weaker condition than $\delta > 2/(\zeta - 2)$ which is required by Connor and Korajczyk (1993). Since minimally, $\delta > 1$ by assumption 3 it follows that ζ is at most needed to be equal to 2.

5.2 Proof of Lemma 2

We must prove that the largest and therefore all eigenvalues of $1/T\hat{\epsilon}'\hat{\epsilon}$ are bounded. To see this we first write

$$1/T\hat{\epsilon}'\hat{\epsilon} = 1/T\epsilon'\epsilon + 1/T(\hat{C} - C)'(\hat{C} - C) + 2/T(\hat{C} - C)'\epsilon = A + B + C$$

Then, the result follows if the maximum eigenvalue of each of A, B, C is bounded. The maximum eigenvalue of $1/T\epsilon'\epsilon$ is bounded by assumption 4. By Theorem 3 of Bai (2003)

$$\hat{C} - C = O_p(\min\{N^{-1/2}, T^{-1/2}\}) = O_p(N^{-1/2})$$

since we assume that $\lim_{N,T \rightarrow \infty} N/T = c$. Thus, the diagonal elements of $(\hat{C} - C)'(\hat{C} - C)$ denoted \hat{C}_{ii}^{cc} are $O_p(1)$ and since

$$\sum_{i=1}^N \mu_i^C = \sum_{i=1}^N \hat{C}_{ii}^{cc} = O_p(N)$$

where μ_i^C are the eigenvalues of $(\hat{C} - C)'(\hat{C} - C)$ it follows that $\mu_1^C = O_p(N)$ and therefore $1/T\mu_1^C = O_p(1)$. Similarly, the i, j -th element of $(\hat{C} - C)\epsilon$ is given by $\hat{C}_{ij}^{cc} = \sum_{s=1}^T \hat{C}_{si}\epsilon_{i,s}$

where \hat{C}_{ij} is the i, j -th element of $\hat{C} - C$. We need to examine \hat{C}_{ij} . By the proof of theorem 3 of Bai (2003) we see that

$$\sqrt{N}\hat{C}_{ij} = \lambda'_i \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{k=1}^N \lambda_k \epsilon_{k,t} + 1/c f_t \left(\frac{F'F}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{s=1}^T f_s \epsilon_{i,s} + o_p(1)$$

Given that \hat{C}_{ij}^{ce} is defined as a sum across t , it is useful to note that $\sqrt{N}\hat{C}_{ij}$ can be written as $\xi_{Ni} + \psi_T f_t$ where

$$\xi_{Ni} = \lambda'_i \left(\frac{\Lambda'\Lambda}{N} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{k=1}^N \lambda_k \epsilon_{k,t}$$

and

$$\psi_T = 1/c \left(\frac{F'F}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{s=1}^T f_s \epsilon_{i,s}$$

As a result viewed as a time series $\sqrt{N}\hat{C}_{ij}$ inherits the dependence properties of f_t . Then, a central limit theorem applied to $\sqrt{N}\hat{C}_{ij}$ gives that $\hat{C}_{ij}^{ce} = O_p(1)$. By a similar treatment to that used to derive the order in probability for μ_1^C we get that $\mu_1^{C\epsilon} = O_p(N)$ where $\mu_i^{C\epsilon}$ are the eigenvalues of $(\hat{C} - C)'\epsilon$ and, therefore, that $1/T\mu_1^{C\epsilon} = O_p(1)$ which completes the proof.

Proof of Theorem 1

Define

$$J_{T(N),N,i}(x, P) = Pr_P \left\{ \tau_{T(N),N}(\hat{\mu}_{1+i} - \hat{\mu}_{r^{max}+1}) \leq x \right\}, \quad i = 0, \dots, r^{max} - 1 \quad (8)$$

By assumption 4, $J_{T(N),N,i}(x, P) \rightarrow J(x, P)$ as $N \rightarrow \infty$. The subsampling approximation to $J_{T(N),N,i}(x, P)$ is given by

$$L_{T(b),b,i}(x) = \frac{1}{N_b} \sum_{j=1}^{N_b} 1 \left\{ \tau_{T(b),b}(\hat{\mu}_{1+i}^j - \hat{\mu}_{r^{max}+1}^j) \leq x \right\}, \quad i = 0, \dots, r^{max} - 1 \quad (9)$$

For x_α , where $J(x_\alpha, P) = \alpha$, we need to prove that $L_{T(b),b,i}(x_\alpha) \rightarrow J(x_\alpha, P)$ for the theorem to hold. But, $E(L_{T(b),b,i}(x_\alpha)) = J_{T(b),b,i}(x_\alpha, P)$. Hence, it suffices to show that $Var(L_{T(b),b,i}(x_\alpha)) \rightarrow 0$ as $N \rightarrow \infty$. The subsampling approach we use samples without replacement b units of length $T(b)$ starting at some random $t < T(N) - T(b)$ out of the N available units where the time series ordering is not tampered with. Let

$$1_{b,j,i} = 1 \left\{ \tau_{T(b),b}(\hat{\mu}_{1+i}^j - \hat{\mu}_{r^{max}+1}^j) \leq x_\alpha \right\} \quad (10)$$

$$v_{N_b,h,i} = \frac{1}{N_b} \sum_{j=1}^{N_b} Cov(1_{b,j,i}, 1_{b,j+h,i}) \quad (11)$$

Then,

$$\begin{aligned} \text{Var} (L_{T(b),b,i}(x_\alpha)) &= \frac{1}{N_b} \left(v_{N_b,0,i} + 2 \sum_{h=1}^{N_b-1} v_{N_b,h,i} \right) = \\ & \frac{1}{N_b} \left(v_{N_b,0,i} + 2 \sum_{h=1}^{b-1} v_{N_b,h,i} \right) + \frac{2}{N_b} \sum_{h=b}^{N_b-1} v_{N_b,h,i} = V_1 + V_2 \end{aligned} \quad (12)$$

But $V_1 = O(b/N_b) = o(1)$. We next examine V_2 . For this we need to note that $V_2 \leq \tilde{V}_2$ where \tilde{V}_2 is made up of covariances obtained by sampling a contiguous block of N_b series of time series dimension equal to $T(b)$ from the unknown ordering of the elements of $\{\{y_{i,t}\}_{t=1}^\infty\}_{i=1}^\infty$ given by

$$\tilde{y} = \{ \{y_{i_1,t}\}_{t=t_{i_1}^{t_{i_1}+T(b)}}, \dots, \{y_{i_{N_b},t}\}_{t=t_{i_{N_b}}^{t_{i_{N_b}}+T(b)}} \}$$

such that \tilde{y} is a cross sectionally mixing process. The mixing nature of \tilde{y} arises from the fact that there are two unknown orderings (not necessarily the same but assumed the same with loss of generality) such that

$$\tilde{\epsilon} = \{ \{\epsilon_{i_1,t}\}_{t=t_{i_1}^{t_{i_1}+T(b)}}, \dots, \{\epsilon_{i_{N_b},t}\}_{t=t_{i_{N_b}}^{t_{i_{N_b}}+T(b)}} \}$$

and

$$\tilde{\lambda}^f = \{ \{\lambda'_{i_1} f_t\}_{t=t_{i_1}^{t_{i_1}+T(b)}}, \dots, \{\lambda'_{i_{N_b}} f_t\}_{t=t_{i_{N_b}}^{t_{i_{N_b}}+T(b)}} \}$$

are cross sectionally mixing processes. $\tilde{\lambda}^f$ is mixing since there is an ordering that makes the sequence $\lambda_1, \dots, \lambda_N$ cross-sectionally mixing and $\{f_t\}_{t=t_i}^{t_i+T(b)}$ is simply a set of constants when sampling across cross-sections. Tildes will be used below to denote that the unknown mixing ordering is used. But, by Lemma 1 it then follows that $\tilde{v}_{N_b,h} = o(1)$. Hence

$$\tilde{V}_2 = \frac{2}{N_b} \sum_{h=b}^{N_b-1} \tilde{v}_{N_b,h} = o(1)$$

proving the convergence of $L_{T(b),b,i}(x_\alpha)$ to $J(x_\alpha, P)$. In order to complete the proof we finally need to show that

$$\hat{L}_{T(b),b,i}(x_\alpha) = \frac{1}{B} \sum_{j=1}^B 1 \{ \tau_{T(b),b}(\hat{\mu}_{1+i}^j - \hat{\mu}_{r^{max}+1}^j) \leq x \}, \quad i = 0, \dots, r^{max} - 1$$

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

Proof of Theorem 2

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

$$Pr_P (\tau_{T(N),N}(\hat{\mu}_{r+1} - \hat{\mu}_{r^{max}+1}) > \hat{q}_{r,\alpha,N}^b) = o(1) \quad (13)$$

where \hat{q}_{i,α_N}^b solves¹

$$L_{T(b),b,i}(x) = \alpha_N \quad (14)$$

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

$$\lim_{N \rightarrow \infty} Pr_P (\tau_{T(N),N} (\hat{\mu}_{r+1} - \hat{\mu}_{r^{max}+1}) > \hat{q}_{r,\alpha_N}^b) = 1 \quad (15)$$

(13) simply follows by (14) and $\alpha_N \rightarrow 0$. We now have to show that (15) holds. We establish the following three facts. Firstly, by Weyl's Theorem (see, e.g. Lutkepohl (1996, 5.3.2(9))), $\hat{\mu}_{1+r} \geq \bar{\mu}_{1+r}$, $r = 0, \dots, r^{max} - 1$ where $\bar{\mu}_{1+r}$ is the $1+r$ 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. Thirdly, by virtue of the fact that $b = o(N)$, $\hat{q}_{\alpha_N}^b$ is $o_p(\tau_{T(N),N} N)$. The first two facts imply that $\tau_{T(N),N} (\hat{\mu}_{1+i} - \hat{\mu}_{r^{max}+1})$ tends to infinity under $H_{1,r}$ at rate $\tau_{T(N),N} N$. This together with the third fact imply (15).

Proof of Theorem 3

It is sufficient to show the following: Firstly

$$\tau_{N,T}(\hat{\mu}_i - \hat{\mu}_{r^{max}+1}) - \hat{\tau}_{N,T}(\hat{\mu}_i - \hat{\mu}_{r^{max}+1}) = o_p(1) \quad (16)$$

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

$$\hat{\beta} - \beta = o_p(\ln N) \quad (17)$$

But, by (7) we get that $\hat{\beta} - \beta = o_p\left(\log\left(\frac{b_1}{b_2}\right)^{-1}\right)$ and since $\log\left(\frac{b_1}{b_2}\right)^{-1} = o(\log N)$ we get (17).

To get (16) we simply note that it is sufficient to show that

$$\hat{\tau}_{N,T}/\tau_{N,T} = N^{\hat{\beta}}/N^{\beta} = 1 + o_p(1) \quad (18)$$

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

¹In the case where (14) has a continuum of solutions we choose the minimum value of x such that (14) holds.

References

- BAI, J. (2003): “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71, 135–173.
- (2004): “Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data,” *Journal of Econometrics*, Forthcoming.
- BAI, J., AND S. NG (2002): “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221.
- BAI, Z. D., AND J. W. SILVERSTEIN (1998): “No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices,” *Annals of Probability*, 26(1).
- CAMBA-MENDEZ, G., AND G. KAPETANIOS (2005): “Estimating the rank of the Spectral Density Matrix,” *Journal of Time Series Analysis*, 26(1), 37–48.
- CAMBA-MENDEZ, G., G. KAPETANIOS, R. J. SMITH, AND M. R. WEALE (2003): “Tests of rank in Reduced rank regression models,” *Journal of Business and Economic Statistics*, 21(1), 145–155.
- CHAMBERLAIN, G., AND M. ROTHSCILD (1983): “Arbitrage, Factor Structure and Man-Variance Analysis in Large Asset Markets,” *Econometrica*, 51, 1305–1324.
- CONNOR, G., AND R. A. KORAJCZYK (1993): “A Test for the Number of Factors in an Approximate Factor Model,” *Journal of Finance*, 48, 1263–1292.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press.
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): “The Generalised Factor Model: Identification and Estimation,” *Review of Economics and Statistics*, 82, 540–554.
- (2004): “The Generalised Factor Model: Consistency and Rates,” *Journal of Econometrics*, Forthcoming.
- HACHEM, W., P. LOUBATON, AND J. NAJIM (2005a): “The Empirical Eigenvalue distribution of a Gram Matrix: From Independence to Stationarity,” *Mimeo*, <http://arxiv.org/abs/math.PR/0502535>.
- (2005b): “The Empirical Eigenvalue distribution of a Gram Matrix with a Given Variance Profile,” *Mimeo*, <http://arxiv.org/abs/math.PR/0411333>.

- JOHNSTONE, I. M. (2001): “On the Distribution of the Largest Eigenvalue in Principal Component Analysis,” *Annals of Statistics*, 29, 295–327.
- KAPETANIOS, G. (2004): “A New Method for Determining the Number of Factors in Factor Models with Large Datasets,” *Queen Mary University of London Working Paper No. 525*.
- LEWBEL, A. (1991): “The Rank of Demand Systems: Theory and Nonparametric Estimation,” *Econometrica*, 59, 711–730.
- LUTKEPOHL, H. (1996): *Handbook of Matrices*. Wiley.
- MAHALANOBIS, P. (1946): “Sample Surveys of Crop Yields in India,” *Sankya, Series A*, 7, 269–280.
- POLITIS, D. N., AND J. P. ROMANO (1994): “Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions,” *Annals of Statistics*, 22, 2031–2050.
- POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. Springer Verlag.
- ROMANO, J. P. (1989): “Bootstrap and Randomization Tests of some Nonparametric Hypotheses,” *Annals of Statistics*, 17, 141–159.
- SCHWARZ, H., H. R. RUTISHAUSER, AND E. STIEFEL (1973): *Numerical Analysis of Symmetric Matrices*. Prentice Hall.
- STOCK, J. H., AND M. W. WATSON (1989): “New Indices of Coincident and Leading Indicators,” in *NBER Macroeconomics Annual 1989*, ed. by O. J. Blanchard, and S. Fischer. Cambridge, M.I.T. Press.
- (2002): “Macroeconomic Forecasting Using Diffusion Indices,” *Journal of Business and Economic Statistics*, 20, 147–162.
- TRACY, C. H., AND H. WIDOM (1996): “On Orthogonal and Symplectic Matrix Ensembles,” *Communications in Mathematical Physics*, 177, 727–754.
- YIN, Y. Q., Z. D. BAI, AND P. R. KRISHNAIAH (1988): “On the Limit of the Largest Eigenvalue of the Large Dimensional Sample Covariance Matrix,” *Probability Theory and Related Fields*, 78.

Table 1. Experiments A									
N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3	ME
$\theta = 1$									
50	50	4.287	4.687	4.032	8.000	4.000	3.995	7.999	4.000
100	50	4.216	4.000	4.000	5.232	4.000	4.000	4.002	4.000
200	50	4.173	4.000	4.000	4.000	4.000	4.000	4.000	4.009
50	100	4.326	4.055	4.003	6.157	4.000	4.000	4.155	4.000
100	100	4.275	4.000	4.000	7.026	4.000	4.000	5.223	4.000
200	100	4.241	4.000	4.000	4.000	4.000	4.000	4.000	4.000
50	200	4.407	4.004	4.000	4.196	4.000	4.000	4.000	4.000
100	200	4.404	4.000	4.000	4.005	4.000	4.000	4.000	4.000
200	200	4.326	4.000	4.000	4.094	4.000	4.000	4.000	4.000
$\theta = 10$									
50	50	3.477	3.075	1.922	7.970	0.088	0.001	7.040	0.800
100	50	4.389	2.575	1.934	4.049	0.133	0.019	2.098	2.945
200	50	4.458	2.515	2.158	3.442	0.236	0.085	1.284	3.972
50	100	4.183	2.558	1.933	4.072	0.132	0.015	2.093	1.563
100	100	4.353	2.782	1.787	4.745	0.663	0.053	3.978	3.816
200	100	4.274	3.289	2.725	3.997	1.739	0.851	3.907	3.999
50	200	4.231	2.451	2.087	3.353	0.210	0.078	1.282	2.553
100	200	4.273	3.260	2.718	3.996	1.755	0.826	3.908	3.996
200	200	4.265	3.963	3.633	4.000	3.745	2.765	4.000	4.000
$\theta = 19$									
50	50	1.612	1.906	0.534	7.933	0.000	0.000	4.687	0.020
100	50	3.228	0.683	0.208	3.221	0.000	0.000	0.087	0.441
200	50	4.453	0.247	0.103	1.206	0.000	0.000	0.001	2.372
50	100	2.755	0.677	0.168	3.075	0.000	0.000	0.068	0.025
100	100	4.460	0.332	0.011	4.380	0.000	0.000	2.360	1.041
200	100	4.732	0.282	0.046	2.481	0.000	0.000	0.634	3.479
50	200	3.895	0.222	0.094	1.113	0.000	0.000	0.002	0.092
100	200	4.512	0.292	0.070	2.460	0.001	0.000	0.648	2.044
200	200	4.478	0.815	0.113	4.000	0.028	0.000	3.936	3.983
Notes: 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)									
ME refers to the method defined in Algorithm 1 of Kapetanios (2004)									

Table 2. Experiments B									
N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3	ME
$\theta = 1$									
50	50	4.574	7.138	5.751	8.000	4.745	3.972	8.000	7.094
100	50	4.513	6.784	5.841	8.000	4.447	4.044	8.000	8.000
200	50	4.451	6.622	5.994	7.956	4.217	4.055	7.650	8.000
50	100	4.363	5.402	4.648	7.954	4.026	4.000	7.581	4.830
100	100	4.395	4.629	4.022	8.000	4.003	4.000	8.000	7.971
200	100	4.347	4.204	4.005	7.896	4.000	4.000	7.495	8.000
50	200	4.420	4.136	4.034	5.304	4.000	4.000	4.019	4.019
100	200	4.412	4.000	4.000	6.003	4.000	4.000	4.313	5.291
200	200	4.396	4.000	4.000	8.000	4.000	4.000	8.000	8.000
$\theta = 10$									
50	50	2.026	5.741	3.888	8.000	0.132	0.002	8.000	2.273
100	50	2.861	5.357	4.133	7.988	0.156	0.025	7.837	7.996
200	50	3.584	5.147	4.351	7.343	0.232	0.101	2.056	8.000
50	100	3.124	3.474	2.465	7.009	0.037	0.002	2.631	1.616
100	100	4.835	3.013	1.701	8.000	0.207	0.014	8.000	7.715
200	100	5.502	3.143	2.471	7.254	0.678	0.203	5.197	8.000
50	200	4.304	2.199	1.777	3.435	0.019	0.004	0.438	1.867
100	200	5.191	2.657	1.979	4.495	0.450	0.107	3.478	4.736
200	200	5.191	3.604	2.751	8.000	2.342	0.812	7.993	8.000
$\theta = 19$									
50	50	1.088	5.432	3.559	8.000	0.030	0.000	8.000	1.535
100	50	1.470	4.873	3.653	7.939	0.032	0.000	7.088	7.938
200	50	1.608	4.640	3.899	6.868	0.016	0.002	0.852	8.000
50	100	1.323	2.595	1.453	6.470	0.000	0.000	0.602	0.334
100	100	2.030	1.663	0.278	8.000	0.000	0.000	8.000	6.035
200	100	2.568	1.194	0.406	6.378	0.000	0.000	1.706	8.000
50	200	2.263	0.441	0.158	1.995	0.000	0.000	0.000	0.064
100	200	3.945	0.197	0.032	3.291	0.000	0.000	0.312	2.323
200	200	5.310	0.450	0.016	7.995	0.000	0.000	7.971	8.000
Notes: 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)									
ME refers to the method defined in Algorithm 1 of Kapetanios (2004)									

Table 3. Experiments C									
N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3	ME
$\theta = 1$									
50	50	3.056	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	50	4.699	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	50	6.767	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.438	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	100	5.604	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	100	6.995	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	200	3.276	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	200	4.880	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	200	6.676	8.000	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 10$									
50	50	1.613	7.999	7.853	8.000	7.993	7.285	8.000	8.000
100	50	3.123	8.000	7.999	8.000	8.000	7.974	8.000	8.000
200	50	4.713	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.229	8.000	7.999	8.000	8.000	7.989	8.000	8.000
100	100	5.167	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.050	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	200	4.021	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	200	5.772	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.326	8.000	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 19$									
50	50	1.576	7.996	7.842	8.000	7.988	7.355	8.000	8.000
100	50	2.731	8.000	7.996	8.000	8.000	7.975	8.000	8.000
200	50	4.118	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.219	8.000	8.000	8.000	8.000	7.995	8.000	8.000
100	100	4.756	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	100	6.977	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	200	3.845	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	200	5.453	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.290	8.000	8.000	8.000	8.000	8.000	8.000	8.000
Notes: 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)									
ME refers to the method defined in Algorithm 1 of Kapetanios (2004)									

Table 4. Experiments D									
N	T	MED	PC_1	PC_2	PC_3	IC_1	IC_2	IC_3	ME
$\theta = 1$									
50	50	2.731	8.000	8.000	8.000	8.000	7.999	8.000	8.000
100	50	4.932	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	50	6.918	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.244	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	100	5.559	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.052	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	200	3.568	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	200	5.116	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	200	6.901	8.000	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 10$									
50	50	1.761	8.000	7.853	8.000	7.997	7.329	8.000	8.000
100	50	2.919	8.000	7.998	8.000	8.000	7.966	8.000	8.000
200	50	4.443	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.237	8.000	8.000	8.000	8.000	7.992	8.000	8.000
100	100	5.109	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	100	7.026	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	200	4.066	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	200	5.751	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.340	8.000	8.000	8.000	8.000	8.000	8.000	8.000
$\theta = 19$									
50	50	1.576	7.999	7.862	8.000	7.993	7.391	8.000	8.000
100	50	2.791	8.000	7.996	8.000	8.000	7.971	8.000	8.000
200	50	4.192	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	100	3.234	8.000	8.000	8.000	8.000	7.996	8.000	8.000
100	100	4.973	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	100	6.833	8.000	8.000	8.000	8.000	8.000	8.000	8.000
50	200	4.109	8.000	8.000	8.000	8.000	8.000	8.000	8.000
100	200	5.433	8.000	8.000	8.000	8.000	8.000	8.000	8.000
200	200	7.240	8.000	8.000	8.000	8.000	8.000	8.000	8.000
Notes: 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)									
ME refers to the method defined in Algorithm 1 of Kapetanios (2004)									

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2005 George Kapetanios
All rights reserved**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**