Inference on the number of factors in factor models

Lorenzo Trapani*

Cass Business School

January 12, 2015

Abstract

This paper proposes a procedure to estimate the number of common factors k in a static approximate factor model. Similarly to other contributions in the literature, the building block of the analysis is the fact that the first k eigenvalues of the covariance matrix of the data diverges, whilst the other stay bounded. On the grounds of this, we firstly propose a test for the null that the i-th eigenvalue diverges, and then employ the tests in a sequential manner to determine k. Based on Monte Carlo evidence, the tests have the correct size and very good power.

JEL codes: C13, C33.

Keywords: approximate factor models, randomised tests, number of factors.

*Cass Business School, Faculty of Finance, 106 Bunhill Row, London EC1Y 8TZ. Tel.: +44 (0) 207 0405260; email: L.Trapani@city.ac.uk.

1 Introduction

This paper proposes a procedure to determine the number of factors (henceforth denoted as k) in a static approximate factor model, viz.

$$X_{it} = \phi_i' F_t + u_{it}, \tag{1}$$

for $1 \leq i \leq N$, $1 \leq t \leq T$, and ϕ_i and F_t are column vectors of dimension k with $0 \leq k \leq \min\{N, T\}$.

Starting from the seminal contribution by Chamberlain and Rothschild (1983), equation (1) has been the subject of several studies; especially in recent years, many contributions have focused on inference on (1), particularly in the context of panel data where both N and T are large - see the review of Bai and Ng (2008), among other contributions, for a comprehensive presentation of the inferential theory for (1).

The first step in the analysis of (1) is, arguably, the determination of the number of common factors k. The literature has developed procedures to determine the number of common factors in both static factor models (see e.g. Bai and Ng, 2002; Alessi, Barigozzi and Capasso, 2009; Onatski, 2010; and Ahn and Horenstein, 2013) and dynamic factor models (e.g. Forni, Hallin, Lippi and Reichlin, 2000; Hallin and Liska, 2007; Amengual and Watson, 2007; and Onatski, 2009). Broadly speaking, all the extant contributions can be grouped into two methodologies: one based on information criteria (IC), and one based on testing. Particularly, IC-based estimation minimizes some penalty function with respect to the number of factors k, over a set defined as $1 \leq k \leq k_{\max}$. The choice of the upper bound $k_{\max}$ is typically done *a priori* (as is the case e.g. in Bai and Ng, 2002, and in Alessi, Barigozzi and Capasso, 2010), and it can have an impact on the estimator of k (see the simulations in Ahn and Horenstein, 2013). Furthermore, the choice of the penalty function is not unique (see Hallin and Liska, 2007), and although refinements have been developed (e.g. Alessi, Barigozzi and Capasso, 2010), this is bound to affect at least the finite sample properties of the estimated k. On the other hand, the latter methodology (based on choosing k on the grounds of some testing procedure) requires to control the procedure-wise rejection rates, and in this respect the literature has not been yet fully developed.

This paper belongs in the latter category, and it is based on the well-known fact that in a panel factor model with k common factors the first k eigenvalues of the covariance matrix of the observations diverge (whilst the other eigenvalues are bounded). Formally, defining $X_t \equiv$

$[X_{1t}, ..., X_{Nt}]'$, we propose a test for the null that the i-th eigenvalue (say $\lambda^{(i)}$) of the covariance matrix $E(X_t X_t')$ diverges as $(N, T) \to \infty$, versus the alternative that it is bounded, viz.

$$\begin{cases} H_0 : \lambda^{(i)} \to \infty \\ H_A : \lambda^{(i)} < \infty \end{cases} . \tag{2}$$

The tests, for $1 \leq i \leq N$, are then employed as part of a sequential procedure to determine k.

From a methodological point of view, the test statistics employed in this paper mimic the behaviour of $\lambda^{(i)}$ - that is they diverge to positive infinity under H_0. Due to the lack of randomness under the null, we base our tests on randomising the test statistic. This approach is not new, *per se*, in the literature. Whilst the original idea dates back to Pearson (1950), it has been recently popularised in econometrics by Corradi and Swanson (2002, 2006) and Bandi and Corradi (2014), *inter alia*. In particular, in this paper we follow the approach used in Corradi and Swanson (2006), where randomisation is employed in conjunction with sample conditioning: randomness is added to the basic statistic, and then the asymptotics is derived conditional on the sample, showing its validity for all samples save for a set of zero measure. Evidence from synthetic data shows that the tests have the correct size and very good power properties; when employing the tests in a sequential fashion to estimate the number of common factors k, results are, again, very good even for small sample sizes.

The remainder of the paper is organised as follows. In Section 2, we discuss the test, its theoretical properties (null distribution and consistency), and possible extensions. Section 4 contains a set of simulations to verify the properties of individual tests, and of the whole procedure to determine k. Section 5 concludes. Proofs are in Appendix.

NOTATION We denote the ordinary limits as "$\to$"; convergence in distribution as "$\xrightarrow{d}$"; convergence in probability and almost surely as "$\xrightarrow{p}$" and "$\xrightarrow{a.s.}$" respectively. We use "a.s." as short-hand for "almost surely", and "$\equiv$" for definitional equality; for a vector a, $\|a\|$ denotes the Euclidean norm. Finite constants that do not depend on the sample size are denoted as M, M', ..., etc. Other relevant notation is introduced in the remainder of the paper.

2 Preliminary theory and assumptions

To being with, note that equation (1) can also be written in matrix form as

$$X_t = \Phi F_t + u_t; \tag{3}$$

in (3), we define $u_t \equiv [u_{1t}, ..., u_{nt}]'$ and the matrix Φ is defined as $\Phi \equiv [\phi_1, ..., \phi_n]'$. Assuming, without loss of generality, that $E(X_{it}) = 0$ for all (i, t), and that factors and idiosyncratic errors are orthogonal, the $N \times N$ covariance matrix of X_t is given by

$$E(X_t X_t') \equiv \Sigma_X = \Phi \Sigma_F \Phi' + \Sigma_u, \tag{4}$$

where $\Sigma_F \equiv E(F_t F_t')$ and $\Sigma_u \equiv E(u_t u_t')$.

The following notation will be used extensively henceforth. We define: the i-th eigenvalue (sorted in descending order) of Σ_X as $\lambda^{(i)}$; the i-th eigenvalue of $\Phi \Sigma_F \Phi'$ as $\gamma^{(i)}$; and, finally, the i-th eigenvalue of Σ_u as $\omega^{(i)}$.

Consider the following assumption.

Assumption 1. It holds that *(i)* $\gamma^{(i)} = m_i N$ for $1 \leq i \leq k$ and some $m_i > 0$; *(ii)* $\omega^{(i)} \leq M < \infty$ for all $1 \leq i \leq N$; *(iii)* $N^{-1} \sum_{i=1}^{N} \gamma^{(i)} \leq M < \infty$ for all N.

By construction, $\gamma^{(i)} = 0$ for $k + 1 \leq i \leq N$. Assumption 1 adds some structure to the rest of the spectra of $\Phi \Sigma_F \Phi'$ and Σ_u. As far as the $\omega^{(i)}$s are concerned, we require that they all be finite; however, they do not need to be bounded away from zero, and some or all of them could indeed be zero. As far as the non-zero $\gamma^{(i)}$s are concerned, part *(i)* of the assumption requires that they diverge to positive infinity, as $N \to \infty$, at a rate $O(N)$. This result is typical of factor analysis: as an example, Bai and Ng (2002) require that, in addition to Σ_F being positive definite, $N^{-1}\Phi'\Phi$ tends to a positive definite matrix, which is tantamount to assuming that $\gamma^{(i)}$ passes to infinity at a rate $O(N)$.

Indeed, there is no reason why the rate of divergence ought to be restricted to $O(N)$. In Section 3.3.1, we consider the more general case of $\gamma^{(i)} = m_i N^{1-\nu_i}$ with $\nu_i \in [0, 1)$, studying in which cases the test manages to determine that the i-th eigenvalue of Σ_X is diverging.

The following well-known result characterizes the eigenvalues of Σ_X.

Lemma 1 *Let $c^{(i)}$, $1 \leq i \leq N$, be a sequence of nonnegative finite numbers, which are strictly positive for $i \leq k$. Then, under Assumption 1(i)-(ii), it holds that, as $N \rightarrow \infty$*

$$\left\{ \begin{array}{c} \frac{\lambda^{(i)}}{N} \rightarrow c^{(i)} \text{ for } 1 \leq i \leq k \\ \lambda^{(i)} \rightarrow c^{(i)} \text{ for } k+1 \leq i \leq N \end{array} \right. . \tag{5}$$

Further, define

$$\bar{\lambda}_N \equiv \frac{1}{N} \sum_{i=1}^{N} \lambda^{(i)}; \tag{6}$$

under Assumption 1(i)-(iii), it holds that

$$\left\{ \begin{array}{c} \limsup_{N \rightarrow \infty} \bar{\lambda}_N = \bar{\lambda}^{\sup} < \infty \\ \liminf_{N \rightarrow \infty} \bar{\lambda}_N = \bar{\lambda}^{\inf} > 0 \end{array} \right. . \tag{7}$$

According to Lemma 1, $\lambda^{(i)}$ either diverges at a rate $O(N)$, or it converges to a finite constant according as $i \leq k$ or not. Basically, the behaviour of the eigenvalues of Σ_X as N passes to infinity is the same as that of the eigenvalues of $\Phi \Sigma_F \Phi'$.

2.1 Estimation of $\lambda^{(i)}$

Let the eigenvector associated with the i-th eigenvalue be denoted by $x^{(i)}$, with the restriction $x^{(i)\prime} x^{(j)} = 1$ or 0 according as $i = j$ or not; the couple $\left(\lambda^{(i)}, x^{(i)}\right)$ is defined by the relationship

$$\Sigma_X x^{(i)} = \lambda^{(i)} x^{(i)}. \tag{8}$$

Consider now the following estimator of Σ_X

$$\hat{\Sigma}_X \equiv \frac{1}{T} \sum_{t=1}^{T} X_t X_t',$$

and let $\left(\hat{\lambda}^{(i)}, \hat{x}^{(i)}\right)$ denote the i-th couple eigenvalue-eigenvector of $\hat{\Sigma}_X$, with the $\hat{x}^{(i)}$s restricted to be an orthonormal basis. The estimator of $\lambda^{(i)}$ satisfies the following equation

$$\hat{\Sigma}_X \hat{x}^{(i)} = \hat{\lambda}^{(i)} \hat{x}^{(i)}. \tag{9}$$

In order to discuss the asymptotic properties of $\hat{\lambda}^{(i)}$, define the (h, j)-th element of Σ_X as

$E\left(X_{ht}X_{jt}\right) \equiv \sigma_{hj}$, and consider the following assumption.

Assumption 2. It holds that $E\left[\max_{1\leq\tilde{t}\leq T}\left|\sum_{t=1}^{\tilde{t}}\left(X_{ht}X_{jt}-\sigma_{hj}\right)\right|^{2}\right] \leq MT$ for $1 \leq h, j \leq N$.

Assumption 2 is akin to Rosenthal-type inequalities (see Rosenthal, 1970; and also Merlevede and Peligrad, 2013), and it could be shown to hold under more primitive assumptions on serial dependence. For example, if the process $X_{ht}X_{jt} - \sigma_{hj}$ were an MDS, such that $E\left|X_{ht}X_{jt}\right|^{2} \leq M$, then one could use, sequentially, Doob's maximal inequality and Burkhloder's inequality, obtaining the bound. Similar results could be obtained assuming that the X_{ht}s are strong mixing: in this case, $X_{ht}X_{jt}$ also is strong mixing (see e.g. Theorem 14.1 in Davidson, 2002), and Rio's (1995) maximal inequality provides the desired bound. In addition to implicitly assuming something about the presence and extent of serial dependence, Assumption 2 also contains a fourth-moment condition. Indeed, it can be expected that $E\left|X_{it}\right|^{4} < \infty$ is needed in order for it to hold. Note that in general the literature considers moment conditions, and dependence conditions, on F_t and u_{it} separately - see e.g. Bai and Ng (2002) and Onatski (2009, 2010).

We now present the asymptotics of $\hat{\lambda}^{(i)}$, reporting an a.s. rate of convergence. We define the functions $f_1(u)$ and $f_2(u)$, chosen as quasimonotone non-decreasing, viz.

$$\lim \sup_{u,v\to\infty} \sup_{0\leq x\leq u} \frac{f_1(x)}{f_1(u)} < \infty,$$

and similarly for $f_2(x)$, and such that

$$\int_1^\infty \frac{1}{u\left[f_1(u)\right]^2}du < \infty \text{ and } \int_1^\infty \frac{1}{u\left[f_2(u)\right]^2}du < \infty. \tag{10}$$

The rate of convergence of $\hat{\lambda}^{(i)}$ is reported in the following Lemma.

Lemma 2 *Under Assumption 2, it holds that*

$$\hat{\lambda}^{(i)} = \lambda^{(i)} + O_{a.s.}\left[\frac{N}{\sqrt{T}}f_1^2(N) f_2(T)\right], \tag{11}$$

for $1 \leq i \leq N$.

Lemma 2 contains the a.s. rate of convergence of $\hat{\lambda}^{(i)}$. In equation (11), $f_1(u)$ and $f_2(u)$ can be any slowly varying function as long as (10) holds - e.g. $f_1(u) = f_2(u) = \sqrt{\ln^{1+\varepsilon} u}$, with $\varepsilon > 0$.

3 The test

In this section, we define the test statistic and study its asymptotics under the null and its power.

To begin with, we consider tests for

$$\begin{cases} H_0 : \lambda^{(i)} = m_i N \\ H_A : \lambda^{(i)} = m_i < \infty \end{cases},$$

for some $m_i > 0$ and finite. The more general case of

$$\begin{cases} H_0 : \lambda^{(i)} = m_i N^{1-v_i} \\ H_A : \lambda^{(i)} = m_i < \infty \end{cases},$$

for some $v_i \in [0, 1)$, is discussed in Section 3.3.1.

3.1 The test statistic

Under the working assumption that the dataset is such that $N < T$, define $\beta \in (0, 1)$ as $N = T^{\beta}$, or alternatively

$$\beta \equiv \frac{\ln N}{\ln T}. \tag{12}$$

Define also $\delta < 1$, such that

$$\delta \begin{cases} \geq 1 - \frac{1}{2\beta} & \text{according as} & \beta \in \left[\frac{1}{2}, 1\right) \\ \geq 0 & & \beta < \frac{1}{2} \end{cases}. \tag{13}$$

The value of δ, if $\beta \geq \frac{1}{2}$, can be chosen arbitrarily, as long as the restriction $\delta \geq 1 - \frac{1}{2\beta}$ is satisfied. Similarly, when $\beta < \frac{1}{2}$, in principle any value of δ can be chosen. Finally, consider the following estimate of $\bar{\lambda}_N$

$$\widehat{\bar{\lambda}}_N \equiv \frac{1}{N} tr\left(\hat{\Sigma}_X\right). \tag{14}$$

We are now ready to present the test. Define the statistic

$$T^{(i)} \equiv N^{-\delta} \frac{\hat{\lambda}^{(i)}}{\widehat{\bar{\lambda}}_N}. \tag{15}$$

Heuristically, it can be expected that, as $(N, T) \to \infty$, $T^{(i)} \to \infty$ at a rate $N^{1-\delta}$ under the null that $\lambda^{(i)} = m_i N$, and that $T^{(i)}$ converges to a finite number under the alternative that $\lambda^{(i)} \leq M$. The denominator $\widehat{\widehat{\lambda}}_N$ is bounded away from zero and infinity on the grounds of (7), and it serves the purpose of normalising the test statistic so as to make it scale-free. Rather than using $T^{(i)}$ directly, we employ the transformation

$$\sqrt{\varphi^{(i)}} \equiv \exp\left\{T^{(i)}\right\}. \tag{16}$$

Under the null that $\lambda^{(i)} = m_i N$, heuristically one can expect that $\sqrt{\varphi^{(i)}} \simeq \exp\left\{\frac{m_i}{\lambda_N} N^{1-\delta}\right\}$; conversely, under the alternative that $\lambda^{(i)} \leq M$ it can be expected that $\sqrt{\varphi^{(i)}} \propto \exp\left\{N^{-\delta}\right\}$, so that

$$\varphi^{(i)} \to \begin{cases} 1 & \text{according as} & \delta > 0 \\ c & & \delta = 0 \end{cases},$$

for some $c \in (0, \infty)$.

Since, under H_0, $\varphi^{(i)} \to \infty$, we do not use it directly, but use a randomised version of it. We present the construction of the test statistic as a four step algorithm.

Step 1 Generate an artificial sample $\left\{\xi_j^{(i)}\right\}_{j=1}^{R}$ as *i.i.d.* $N(0,1)$, and define $\sqrt{\varphi^{(i)}} \times \xi_j^{(i)}$, $1 \leq j \leq R$;

Step 2 Define the sample $\left\{\zeta_j^{(i)}(u)\right\}_{j=1}^{R}$ as

$$\zeta_j^{(i)}(u) \equiv I\left[\sqrt{\varphi_i} \times \xi_j^{(i)} \leq u\right], \tag{17}$$

for some nonzero u, extracted from a distribution $F(u)$ with support $U \subset R \backslash \{0\}$;

Step 3 Compute the normalised sum

$$\vartheta^{(i)}(u) \equiv \frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left[\zeta_j^{(i)}(u) - \frac{1}{2}\right]; \tag{18}$$

Step 4 Define the test statistic

$$\Theta^{(i)} \equiv \int_U \left[\vartheta^{(i)}(u)\right]^2 dF(u). \tag{19}$$

In order to provide a heuristic preview of how the test statistic works, consider (17) and (18). Under the null, φ_i passes to infinity, so that the variance of $\sqrt{\varphi_i} \times \xi_j^{(i)}$ should be ∞; hence, under the null, the *i.i.d.* sequence $\left\{\zeta_j^{(i)}(u)\right\}_{j=1}^{R}$ should follow a Bernoulli distribution with $E\left[\zeta_j^{(i)}(u)\right]=\frac{1}{2}$. Therefore, in (18) a CLT should hold whereby, as $R\rightarrow\infty$, $\vartheta^{(i)}(u)$ should be $N(0,1)$. Conversely, under the alternative, φ_i should remain finite, and therefore it can be expected that, for any $u\neq 0$, $E\left[\zeta_j^{(i)}(u)\right]\neq\frac{1}{2}$. Thus, in (18), there is a sum of *i.i.d.* random variables with nonzero mean: a LLN should entail that $\vartheta^{(i)}(u)$ diverges to positive infinity, at a speed $\sqrt{R}$.

The choice of the two main specifications - the size of the artificial sample R, and the value(s) of u - are discussed after Theorems 1 and 2, and also in Section 4.

3.2 Asymptotic properties

We now discuss the null distribution and the power versus the alternative $H_A:\lambda^{(i)}\leq M<\infty$ for tests based on $\Theta^{(i)}$. As mentioned in the Introduction, we employ a randomised testing approach conditioning on the sample, and showing that our results (null distribution and power) hold for all samples, apart from a set of measure zero.

Henceforth, we frequently employ the following notation. We define P^* as the probability law of $\left\{\zeta_j^{(i)}(u)\right\}_{j=1}^{R}$ conditional on the sample, and we let "$\overset{d^*}{\rightarrow}$" denote convergence in distribution according to P^*.

The following theorem characterizes the null distribution of $\Theta^{(i)}$, and it also provides a selection rule for R.

Theorem 1 *Let Assumptions 1 and 2 hold. Then, under* $H_0:\lambda^{(i)}=m_iN$, *as* $(N,T,R)\rightarrow\infty$ *with* $\frac{N^{1-\delta}}{\sqrt{T}}\rightarrow 0$ *and*

$$\frac{R}{\exp\left\{\epsilon N^{1-\delta}\right\}}\rightarrow 0, \tag{20}$$

for some $0<\epsilon<2\frac{m_i}{\lambda_N}$, *it holds that* $\Theta^{(i)}\overset{d^*}{\rightarrow}\chi_1^2$ *a.s.-*P^* *conditionally on the sample.*

Theorem 1 states that, under the null, the test statistic $\Theta^{(i)}$ follows a chi-squared distribution with one degree of freedom.

In order for Theorem 1 to hold, it is necessary that $R\rightarrow\infty$, which is natural since equation (18) is an application of the CLT. Given that, in (18), the $\zeta_j^{(i)}(u)$s are *i.i.d.* and have a uniform distribution, it can be expected that convergence should be quite fast. As far as selecting R is concerned, equation (20) provides an upper bound. When implementing the test, we set $R=N$; in principle, any choice of $R\rightarrow\infty$ with (20) is acceptable.

We now turn to analysing the consistency of the test versus the alternative $H_A:\lambda^{(i)}\leq M<\infty$.

Theorem 2 *Let Assumptions 1 and 2 hold, and define* c_α *as* $P\left[\Theta^{(i)} \leq c_\alpha\right] = \alpha$ *under* H_0. *Under* H_A, *as* $(N, T, R) \to \infty$ *with* $\frac{N^{1-\delta}}{\sqrt{T}} \to 0$, *it holds that* $P\left[\Theta^{(i)} > c_\alpha\right] = 1$ *a.s.-*P^* *conditionally on the sample.*

In the proofs of Theorems 1 and 2, we show that $\vartheta^{(i)}(u)$ has a non-centrality parameter asymptotically equal to

$$\frac{2}{\sqrt{R}} \sum_{j=1}^{R} \int_{0}^{|u|} \frac{1}{\sqrt{2\pi\varphi^{(i)}}} \exp\left\{-\frac{1}{2}\frac{t^2}{\varphi^{(i)}}\right\} dt = \sqrt{\frac{2R}{\pi}} \left[\frac{|u|}{\sqrt{\varphi^{(i)}}} - \frac{1}{6}b^3\right],$$

where $b \in \left(0, \frac{|u|}{\sqrt{\varphi^{(i)}}}\right)$. Under the null, this term should go to zero; given that $\varphi^{(i)} \simeq \exp\left\{2\frac{m_i}{\lambda_N}N^{1-\delta}\right\}$, this means that

$$\sqrt{\frac{2u^2}{\pi}\frac{R}{\exp\{2N^{1-\delta}\}}} \to 0, \tag{21}$$

whence (20). Under the alternative, the term is bounded from below by

$$\sqrt{\frac{2R}{\pi}}\frac{|u|}{\sqrt{\varphi^{(i)}}}\left[1 - \frac{1}{6}\frac{u^2}{\varphi^{(i)}}\right], \tag{22}$$

and in order to have power this should pass to infinity. Note that the expression above is maximised for $|u| \to \infty$: however, a large value of $|u|$ may yield size distortion in view of (21). Alternatively, the expression (22) has a local maximum at $|u| = \sqrt{2\varphi^{(i)}}$. Note that if $\delta > 0$ in (15), then $\varphi^{(i)}$ converges to 1: these heuristic considerations point to a choice of $u = \sqrt{2}$, or $u = -\sqrt{2}$, or u taking values $\pm\sqrt{2}$ with equal probability. This is further supported by Monte Carlo evidence (Section 4).

3.3 Discussion and extensions

In this Section, we discuss the implementation of the test, and its ability to distinguish between eigenvalues that are finite from eigenvalues that diverge slowly to positive infinity.

3.3.1 Type I error in presence of slowly diverging eigenvalues

Assumption 1 stipulates that the eigenvalues of Σ_X either are bounded or diverge at a rate N. In this section, we study what happens when the eigenvalues diverge at a slower rate than N. This is the case of "weak factors", i.e. the case in which one or more common factor does give a contribution, to the covariance matrix of the data Σ_X, that passes to infinity, but at a rate slower than $O(N)$.

Formally, we study the properties of tests based on $\Theta^{(i)}$ for

$$\begin{cases} H_0 : \lambda^{(i)} = m_i N^{1-\nu_i} \\ H_A : \lambda^{(i)} \leq M < \infty \end{cases}, \tag{23}$$

for some $v_i \in [0, 1)$.

Our finding is that the test is able to accept H_0 in this context, with a probability of a Type I error equal to a given level α (say $\alpha = 0.05$); however, this depends on ν_i, and on the relative rate of divergence between N and T as they pass to infinity. Thus, the tests derived in this paper are able to detect the presence of weak factors, as long as they are not "too weak". As Forni *et al.* (2000, p. 547) put it, "[...] there is no way a slowly diverging sequence (divergence under the model can be arbitrarily slow) can be told from an eventually bounded sequence (for which the bound can be arbitrarily large)".

Consider the following assumption, which extends Assumption 1.

Assumption 3. It holds that *(i)* $\gamma^{(i)} = m_i N^{1-\nu_i}$ with $\nu_i \in [0, 1)$ for $1 \leq i \leq k$ and some $m_i > 0$; *(ii)* $\omega^{(i)} \leq M < \infty$ for all $1 \leq i \leq N$; *(iii)* (a) $N^{-1} \sum_{i=1}^{N} \gamma^{(i)} \leq M < \infty$ for all N, and (b) either $\omega^{(N)} > 0$ or $N^{-1} \sum_{i=1}^{N} \gamma^{(i)} \geq M' > \infty$ for all N.

The assumption is the same as Assumption 1. The only difference is part (b), which implies a lower bound on $\frac{1}{N} \sum_{i=1}^{N} \lambda^{(i)}$. This can be obtained either by assuming that Σ_u has full rank, or that there are many common factors - i.e. that k grows with N.

It holds that

Lemma 3 *Let $c^{(i)}$, $1 \leq i \leq N$, be a sequence of nonnegative finite numbers, which are strictly positive for $i \leq k$. Then, under Assumption 3(i)-(ii), it holds that, as $N \to \infty$*

$$\begin{cases} \frac{\lambda^{(i)}}{N^{1-\nu_i}} \to c^{(i)} \text{ for } 1 \leq i \leq k \\ \lambda^{(i)} \to c^{(i)} \text{ for } k+1 \leq i \leq N \end{cases}. \tag{24}$$

Further, under Assumption 3(i)-(iii), it holds that

$$\begin{cases} \lim_{N\to\infty} \sup \bar{\lambda}_N = \bar{\lambda}^{\sup} < \infty \\ \lim_{N\to\infty} \inf \bar{\lambda}_N = \bar{\lambda}^{\inf} > 0 \end{cases}. \tag{25}$$

We now give a heuristic preview of the arguments that lead to the main result. Theorem 1 states that test is able to accept the null with a Type I error probability of α as long as (20) holds. In view of the proof of Theorem 1, (20) arises from the test statistic having a non-centrality term proportional to $\sqrt{\frac{R}{\varphi^{(i)}}}$; in turn, $\varphi^{(i)}$ diverges to infinity, under the null, at a rate $\exp\left\{2\frac{m_i}{\lambda_N}N^{-\delta}\right\}$. Thus, under (23), Theorem 1 holds if

$$\frac{R}{\exp\left\{N^{\epsilon[1-(\nu_i+\delta)]}\right\}} \to 0,$$

for some $0 < \epsilon < 2\frac{m_i}{\lambda_N}$. This requires that $\nu_i + \delta < 1$. The following corollary (based on the choice $R = N$) formalises the result.

Corollary 1 *Let $\chi^2_{1,1-\alpha}$ be the quantile of level α from the χ^2_1 distribution, and let Assumptions 2 and 3 hold. Under $H_0 : \lambda^{(i)} = m_i N^{1-\nu_i}$, as $(N, T, R) \to \infty$ with $\frac{N^{1-\delta}}{\sqrt{T}} \to 0$, if it holds that*

$$\nu_i + \delta < 1, \tag{26}$$

then it holds that

$$P\left[\Theta^{(i)} \leq \chi^2_{1,1-\alpha}\right] = \alpha \text{ a.s.-}P^*.$$

Further, under $H_A : \lambda^{(i)} \leq M < \infty$, as $(N, R, T) \to \infty$ with (20), it holds that $P\left[\Theta^{(i)} > \chi^2_{1,1-\alpha}\right] = 1$ a.s.-P^*.

Corollary 1 stipulates that tests based on $\Theta^{(i)}$ can discern between diverging from bounded eigenvalues even when eigenvalues do not pass to infinity, under the null, at a rate of strictly $O(N)$. Since the null is that eigenvalues diverge, Corollary 1 does not address the presence of power versus local alternatives, but rather the correct size of the test when the null is "local to the alternative".

The "constructive" part of the corollary is equation (26), which illustrates up to which degree of slowness the test can detect diverging eigenvalues. Consider first the case $\beta < \frac{1}{2}$; by (13), in this case $\delta \geq 0$. Upon choosing $\delta = 0$, (26) boils down to requiring that $\nu_i < 1$: the test has probability of a Type I error equal to α (asymptotically) for *any* polynomial rate of divergence of $\lambda^{(i)}$. Considering the case $\beta \in \left[\frac{1}{2}, 1\right)$, in view of (13), we have $\delta \geq 1 - \frac{1}{2\beta}$, so that (26) becomes

$$\nu_i < \frac{1}{2\beta}.$$

When $\beta > \frac{1}{2}$, it is therefore still possible to detect a non-pervasive factor. However, the larger β the smaller ν_i, and thus the less able the test is to detect such factors. In the extreme case that $\beta = 1$, i.e. when $N = T$, the test would not be able to accept the null that $\lambda^{(i)} = m_i\sqrt{N}$ (or smaller) with probability α. The source of this confusion between null and alternative is the estimation error of $\hat{\lambda}^{(i)}$, which grows with N irrespective of whether $\lambda^{(i)}$ is bounded or diverging; indeed, the larger N relatively to T, the larger such error, so that $\hat{\lambda}^{(i)}$, and thus ultimately the test, is driven by the estimation error, and not by the value of $\lambda^{(i)}$ itself.

3.3.2 Determining the number of factors when $N > T$

All the discussion above postulates that $N < T$. This is primarily the reason why we estimate $\Sigma_X \equiv E\left(X_t X_t'\right)$. However, all the results derived in this paper still hold if $N > T$. The purpose of this Section is to clarify how the test should be employed, and which results to expect. As a general comment, the choice of estimating $E\left(X_t X_t'\right)$ is natural in a context where $N < T$. Should instead one have $N > T$, the results developed in this paper can be applied with no loss of generality, by simply estimating $E\left(X_i X_i'\right)$ - where we have defined $X_i = \left[X_{i1}, ..., X_{iT}\right]'$. All the results derived above hold, simply swapping N and T around.

Model (1) can be rewritten in matrix form as

$$X_i = F\Phi_i + u_i, \tag{27}$$

having defined $X_i = \left[X_{i1}, ..., X_{iT}\right]'$, $F = \left[F_1, ..., F_T\right]'$ and $u_i = \left[u_{i1}, ..., u_{iT}\right]'$. When $N > T$, the most natural covariance matrix to consider, in order to reduce the computational burden of the problem, is

$$\Sigma_X^{(T)} \equiv E\left(X_i X_i'\right) = E\left(F\Phi_i\Phi_i'F'\right) + E\left(u_i u_i'\right) = \Sigma_F^{(T)} + \Sigma_u^{(T)},$$

whose T eigenvalues, sorted in descending order, can be indicated as $\lambda^{(t)}$, $1 \leq t \leq T$. Based on (27), it can be expected that $\hat{\lambda}^{(t)} \to \infty$ at some rate for $1 \leq t \leq k$, whilst $\hat{\lambda}^{(t)} \leq M < \infty$ for $k+1 \leq t \leq T$.

Define now: the eigenvalues of $\Sigma_F^{(T)}$ and $\Sigma_u^{(T)}$ as, respectively, $\gamma^{(t)}$ and $\omega^{(t)}$; the element of $\Sigma_X^{(T)}$ in position (s,t) as σ_{st}. Consider the following assumptions, which mimic Assumptions 1-3 above:

Assumption 1'. It holds that *(i)* $\gamma^{(t)} = m_t N$ for $1 \leq t \leq k$ and some $m_t > 0$; *(ii)* $\omega^{(t)} \leq M < \infty$ for all $1 \leq t \leq T$; *(iii)* (a) $T^{-1}\sum_{i=1}^{T}\gamma^{(t)} \leq M < \infty$ for all T, and (b) either $\omega^{(T)} > 0$ or

$T^{-1}\sum_{i=1}^{T}\gamma^{(t)} \geq M' > \infty$ for all T.

Assumption 2'. It holds that $E\left[\max_{1\leq\tilde{\imath}\leq N}\left|\sum_{i=1}^{\tilde{\imath}}\left(X_{is}X_{it}-\sigma_{st}\right)\right|^2\right] \leq MN$ for $1 \leq s,t \leq T$.

Assumption 3'. It holds that *(i)* $\gamma^{(t)} = m_t N^{1-\nu_t}$ with $\nu_t \in [0,1)$ for $1 \leq t \leq k$ and some $m_t > 0$; *(ii)* $\omega^{(t)} \leq M < \infty$ for all $1 \leq t \leq T$; *(iii)* (a) $T^{-1}\sum_{i=1}^{T}\gamma^{(t)} \leq M < \infty$ for all T, and (b) either $\omega^{(T)} > 0$ or $T^{-1}\sum_{i=1}^{T}\gamma^{(t)} \geq M' > \infty$ for all T.

Testing for

$$\begin{cases} H_0 : \lambda^{(t)} = m_t T \\ H_A : \lambda^{(t)} \leq M < \infty \end{cases}, \tag{28}$$

can be based on estimating $\lambda^{(t)}$ from $\hat{\Sigma}_X^{(T)} = \frac{1}{N}\sum_{i=1}^{N} X_i X_i'$, viz. by using $\left(\hat{\lambda}^{(t)}, \hat{x}^{(t)}\right)$ defined as the solution of

$$\hat{\Sigma}_X^{(T)}\hat{x}^{(t)} = \hat{\lambda}^{(t)}\hat{x}^{(t)}.$$

As before, for each $1 \leq t \leq T$ a test statistic can be defined to test for H_0. Letting $\hat{\bar{\lambda}}_T = \frac{1}{T}tr\left(\hat{\Sigma}_X^{(T)}\right)$, define

$$T^{(t)} \equiv T^{-\delta'}\frac{\hat{\lambda}^{(t)}}{\hat{\bar{\lambda}}_T},$$

where, for $\beta' \equiv \frac{\ln T}{\ln N}$, we set $\delta' \geq 0$ if $\beta' < \frac{1}{2}$ and $\delta' \geq 1 - \frac{1}{2\beta'}$ for $\beta' \in \left[\frac{1}{2},1\right)$. Consider now the transformation $\sqrt{\varphi^{(t)}} = \exp\left\{T^{(t)}\right\}$. Under the null, it can be expected that $T^{(t)} = O\left(N^{1-\delta'}\right)$, whilst $T^{(t)}$ should remain bounded under the alternative. Thence the passages are the same as above: for an artificially generated *i.i.d.* $N(0,1)$ sequence, say $\left\{\xi_j^{(t)}\right\}_{j=1}^{R}$, we define

$$\zeta_j^{(t)}(u) \equiv I\left[\varphi_t^{1/2} \times \xi_j^{(t)} \leq u\right],$$

for some $u \in U\backslash\{0\}$, and

$$\begin{aligned} \vartheta^{(t)}(u) &\equiv \frac{2}{\sqrt{R}}\sum_{j=1}^{R}\left[\zeta_j^{(t)}(u) - \frac{1}{2}\right], \\ \Theta^{(t)} &\equiv \int_U \left[\vartheta^{(t)}(u)\right]^2 f(u)\,du. \end{aligned} \tag{29}$$

The following Lemma (reported without proof) is the counterpart to Theorems 1 and 2.

Lemma 4 *Let Assumptions 1' and 2' hold. Then, under* $H_0 : \lambda^{(t)} = m_t N$*, as* $(N,T,R) \to \infty$

with $\frac{T^{1-\delta'}}{\sqrt{N}} \to 0$ *and*

$$\frac{R}{\exp\left\{\epsilon' T^{1-\delta'}\right\}} \to 0,$$

for some $\epsilon' > 0$*, it holds that* $\Theta^{(t)} \overset{d^*}{\to} \chi^2_1$ *a.s.-*P^* *conditionally on the sample. Defining* c_α *as* $P\left[\Theta^{(t)} \leq c_\alpha\right] = \alpha$ *under* H_0*, as* $(N, T, R) \to \infty$ *and under* H_A*, it holds that* $P\left[\Theta^{(t)} > c_\alpha\right] = 1$ *a.s.-*P^* *conditionally on the sample.*

Finally, let Assumptions 2' and 3' hold. Under $H_0 : \lambda^{(t)} = m_t N^{1-\nu_t}$*, as* $(N, T) \to \infty$*, if* $\nu_t + \delta < 1$*, then it holds that*

$$P\left[\Theta^{(t)} \leq \chi^2_{1,1-\alpha}\right] = \alpha \text{ a.s.-}P^*.$$

Further, under $H_A : \lambda^{(t)} \leq M < \infty$*, as* $(N, R, T) \to \infty$ *with (20), it holds that* $P\left[\Theta^{(t)} > \chi^2_{1,1-\alpha}\right] = 1$ *a.s.-*P^**.*

4 Simulations and empirical exercise

In this section, we evaluate the performance of single tests, and of the sequential procedure to determine k, using synthetic data. The purpose of the former exercise (where individual tests are considered) is primarily to illustrate the finite sample properties of the test. Conversely, the latter exercise (determining k) is aimed at providing some practical guidelines as to how to implement the sequential procedure - e.g. deciding the family-wise rejection rate, the choice of δ defined in (13), the normalisation to be employed in the construction of $T^{(i)}$, etc...

In both exercises, data are generated as

$$X_{it} = \sum_{j=1}^{k} \lambda_{ij} F_{jt} + \sqrt{\theta} u_{it}, \tag{30}$$

where $F_{jt} \sim i.i.d.N(0,1)$ for $1 \leq t \leq T$ and $1 \leq j \leq k$; similarly, $\lambda_{ij} \sim i.i.d.N(0,1)$ for $1 \leq i \leq N$ and $1 \leq j \leq k$. Equation (30) is a very similar design to bai and Ng (2002; see also the simulation section in Ahn and Horenstein, 2013), and the comments that follow are based on their simulation exercise. In (30), the idiosyncratic error u_{it} is generated as

$$u_{it} = \sqrt{\frac{1-\rho^2}{1+2bC}} e_{it}, \tag{31}$$

$$e_{it} = \rho e_{it-1} + v_{it} + b\left(\sum_{h=\max\{i-C,1\}}^{i-1} v_{ht} + \sum_{h=i+1}^{\min\{i+C,N\}} v_{ht}\right). \tag{32}$$

In (32), $v_{it} \sim i.i.d.N(0,1)$ for $1 \leq i \leq N$. The coefficient ρ is used to introduce serial dependence in the error term u_{it}; the component $\sum_{h=\max\{i-C,1\}}^{i-1} bv_{ht} + \sum_{h=i+1}^{\max\{i+C,N\}} bv_{ht}$ in (32) serves the purpose of introducing cross-sectional dependence among the u_{it}s. As in Ahn and Horenstein (2013), in (31) the variance of the u_{it}s is normalised so that for most of them - namely, for the units where $C+1 \leq i \leq N-C$ - it holds that $Var(u_{it}) = 1$.

Hence, the coefficient $\frac{1}{\theta}$ in (30) is equal to the signal-to-noise of the common factors.

In both exercises, data are generated according to four different schemes, which correspond to different levels of serial and cross-sectional correlation:

(a) *i.i.d.* data: corresponding to $\rho = b = C = 0$;

(b) serially correlated, but cross sectionally independent data: $\rho = 0.5$ and $b = C = 0$;

(c) serially independent, but cross-sectionally correlated data: $\rho = 0$, $b = 0.5$ and $C = \max\left\{10, \frac{N}{20}\right\}$;

(d) serially and cross-sectionally correlated data: $\rho = b = 0.5$ and $C = \max\left\{10, \frac{N}{20}\right\}$.

These designs are the same as in Ahn and Horenstein (2013).

Finally, a word on the notion of size in the context of randomised tests: as Corradi and Swanson (2006) point out, a feature of randomisation and sample conditioning is the different interpretation that the notion of test size has in this context. More specifically, in a classical hypothesis testing context, the level α of a test means that, if a researcher applies the test B times and the null is valid, then (s)he will reject the null with frequency α - that is, (s)he will be wrong αB times. Conversely, in our context the level of α is interpreted thus: out of J researchers who apply the test, αJ of them will reject the null when this is true. Despite such interpretational difference, the results in Section 3.2 show that, using the randomised testing approach we obtain a test statistic which, for a given level α, rejects the null with probability α when true, and with probability 1 when false.

4.1 Individual tests: size and power

We consider size and power of tests for the following hypothesis testing framework, based on the second largest eigenvalue of Σ_X

$$\begin{cases} H_0 : \lambda^{(2)} \to \infty \text{ as } (N,T) \to \infty \\ H_0 : \lambda^{(2)} < \infty \text{ as } (N,T) \to \infty \end{cases}.$$

In order to study the size of tests based on $\Theta^{(2)}$, we generate the data with $k = 2$ in (30). Unreported experiments show no differences when considering any other eigenvalues; or when setting $k = 2$ as opposed to $k = 3$. When studying the power, we generate data setting $k = 1$ in (30). As far as the specification of $\Theta^{(2)}$ is concerned, as discussed above we run the test with $R = N$, choosing δ based on (13) as

$$\delta \begin{cases} = 1.2 \times \left(1 - \frac{1}{2\beta}\right) \\ = 0.2 \end{cases} \text{according as} \quad \begin{array}{c} \beta \in \left[\frac{1}{2}, 1\right) \\ \beta < \frac{1}{2} \end{array}. \tag{33}$$

We use $u = \pm\sqrt{2}$ with equal weight. Unreported simulations also show that tests based on $T^{(2)}$ defined in (15) work well in terms of size and power. However, even better results (particularly, as regards the size) are obtained using the modified version

$$\tilde{T}^{(2)} \equiv N^{-\delta} \frac{\hat{\lambda}^{(2)}}{\hat{\bar{\lambda}}_{N,(2)}}, \tag{34}$$

having defined

$$\hat{\bar{\lambda}}_{N,(2)} \equiv \frac{1}{N-2} tr^* \left(\hat{\Sigma}_X\right), \tag{35}$$

with $tr^* \left(\hat{\Sigma}_X\right)$ being computed as the sum of the last $N - 2$ elements, sorted in descending order, on the main diagonal of $\hat{\Sigma}_X$. Intuitively, this still serves the purpose of rendering $\hat{\lambda}^{(2)}$ scale-free, without $\tilde{T}^{(2)}$ being too small. This should improve the size of the test, at the expense of power: simulation results show that the power of the test is anyway very good even after this transformation.

Results, under scenarios *(a)-(d)* for a level $\alpha = 0.05$, are in Table 1. Each Monte Carlo experiment was run with $2,000$ iterations, and hence, when considering the size, the empirical rejection frequencies reported in the table have a confidence interval of $[0.04, 0.06]$.

[Insert Tables 1a-2c somewhere here]

As can be seen, the presence of serial or cross-sectional dependence does not alter the size of the test (compare Table 1a with, respectively, Tables 1b and 1c). The test has good power properties, although the presence of cross dependence (Table 2c) reduces the power when $N \leq 25$; when $N \geq 50$, the impact of cross dependence is absent.

4.2 Determining the number of common factors

As well as using the tests individually, they can be employed sequentially in order to provide an estimate of the number of common factors k. Indeed, this is the main purpose of the theory developed in this paper. In this section, we consider firstly how to implement such sequential procedure, and we then evaluate it performance.

To begin with, estimation of k can be performed using the following algorithm:

Step 1 Run the test for $H_0 : \lambda^{(1)} = \infty$ based on $\Theta^{(1)}$. If the null is rejected, set $\hat{k} = 0$ and stop, otherwise go to the next step;

Step 2 Starting from $i = 1$, run the test for $H_0 : \lambda^{(i+1)} = \infty$ based on $\Theta^{(i+1)}$, constructed using an artificial sample $\left\{\xi_j^{(i+1)}\right\}_{j=1}^{R}$ generated independently of $\left\{\xi_j^{(1)}\right\}_{j=1}^{R}$, ..., $\left\{\xi_j^{(i)}\right\}_{j=1}^{R}$. If the null is rejected, set $\hat{k} = i$ and stop; otherwise set $i = i + 1$ and repeat the step until the null is rejected;

Step 3 The estimate of k is $\hat{k} = i$.

The specification of each test statistic $\Theta^{(i)}$ is based on the same choices as above, namely: $R = N$, δ chosen selected as in (33), and

$$\tilde{T}^{(i)} \equiv N^{-\delta} \frac{\hat{\lambda}^{(i)}}{\hat{\bar{\lambda}}_{N,(i)}},$$

with

$$\hat{\bar{\lambda}}_{N,(i)} \equiv \frac{1}{N-i} tr^* \left(\hat{\Sigma}_X \right),$$

$tr^* \left(\hat{\Sigma}_X \right)$ being computed as the sum of the last $N - i$ elements, sorted in descending order, on the main diagonal of $\hat{\Sigma}_X$.

The only difference between the iterative procedure considered here and the individual test discussed in the previous section is that in this context a sequence of tests is employed, and therefore the question arises as to how to control the family-wise level of the test, given that each individual test has a level α. This can be formalised by considering the following indicator

$$ME = \frac{1}{MC} \sum_{j=1}^{MC} d_j,$$

where MC is the number of iterations in each Monte Carlo experiment, and d_j is a dummy variable defined as 1 if $\hat{k} \neq k$ at iteration j, and zero otherwise. Other indicators could be employed, e.g.

instead of d_j one could employ $\left|\hat{k}_j - k\right|$, $\hat{k}_j$ being the estimate at iteration j - this indicator has indeed been used, but it is not reported since results are very similar to the ones based on ME - this indicates that $\hat{k}_j$ can be wrong, but it is rarely "very wrong". Ideally, it would be desirable to have $ME \leq \alpha$, and as close to zero as possible. In the context of this paper it should be noted that, as long as the artificial samples $\left\{\xi_j^{(i)}\right\}_{j=1}^{R}$ are generated independently for $1 \leq i \leq N$, the test statistics $\Theta^{(i)}$ are mutually independent. Therefore, we employed a standard Bonferroni correction, where each individual test for $H_0 : \lambda^{(i)} = \infty$ is carried out at a level $\frac{\alpha}{N}$. Simulation results are very good in all the cases considered, suggesting that this approach should be employed in applications.

We start by considering the case in which $k = 3$ in (30), all the three factors have the same variance with $Var(F_{jt}) = 1$, and the factors are strong - to this end, we set $\theta = 1$ in (30). We consider the same combinations of N and T as in the previous section, and we set $MC = 1,000$. Results are in Tables 3a-3c:

[Insert Tables 3a-3c somewhere here]

The tables show that the procedure to determine the number of factors always has a percentage of error lower than 5%: this holds true irrespective of the presence and extent of dependence, whether this be serial or cross-sectional. Results are already correct when N is as small as 25, and improve as both N and T increase.

5 Conclusions

This paper proposes a procedure to estimate the number of common factors k in a static approximate factor model. Similarly to other contributions in the literature, the building block of the analysis is the fact that the first k eigenvalues of the covariance matrix of the data diverges, whilst the other stay bounded. On the grounds of this, we firstly propose a test for the null that the i-th eigenvalue diverges, and then employ the tests in a sequential manner to determine k. Based on Monte Carlo evidence, the tests have the correct size and very good power. The procedure to determine k, based on a sequential application of the test, works well even in small samples, with N as small as 25, and is not significantly affected by the presence of either serial or cross-sectional dependence.

Appendix: Proofs and Derivations

Henceforth, we let E^* and V^* denote the expected value and the variance calculated with respect to P^*.

Prior to reporting the proofs of the main results in the paper, consider the following preliminary lemma.

Lemma A.1. *Under Assumptions 1 and 2, it holds that, as* $(N,T)\to\infty$

$$\begin{cases} \lim_{N\to\infty}\sup\widehat{\bar{\lambda}}_N=\bar{\lambda}^{\sup}<\infty \\ \lim_{N\to\infty}\inf\widehat{\bar{\lambda}}_N=\bar{\lambda}^{\inf}>0 \end{cases}.$$

Proof. The proof is similar to that of Lemma 2, and we therefore only report the main passages. Note first that

$$\widehat{\bar{\lambda}}_N-\bar{\lambda}_N\equiv\frac{1}{N}tr\left(\hat{\Sigma}_X\right)-\frac{1}{N}tr\left(\Sigma_X\right)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}^2-\sigma_{ii}\right);$$

let also $\delta_{it}\equiv X_{it}^2-\sigma_{ii}$ for short. We start by showing that

$$\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{NT}P\left[\max_{1\leq\tilde{t}\leq T,1\leq\tilde{i}\leq N}\left|\frac{1}{NT}\sum_{i=1}^{\tilde{i}}\sum_{t=1}^{\tilde{t}}\delta_{it}\right|>\varepsilon f_3(N)f_4(T)\right]<\infty,\tag{36}$$

for some $\varepsilon>0$, where $f_3(N)$ is defined such that it is quasimonotone non-decreasing and

$$\int_1^{\infty}\frac{1}{u\left[f_3(u)\right]^2}du<\infty.$$

Similarly, we define $f_4(T)=\sqrt{T}f_5(T)$, with $f_5(T)$ having the same properties as $f_3(T)$. Consider

$$\begin{aligned} E\left[\max_{1\leq\tilde{t}\leq T,1\leq\tilde{i}\leq N}\left|\frac{1}{NT}\sum_{i=1}^{\tilde{i}}\sum_{t=1}^{\tilde{t}}\delta_{it}\right|^2\right] &\leq E\left[\max_{1\leq\tilde{t}\leq T,1\leq\tilde{i}\leq N}\frac{1}{N}\sum_{i=1}^{\tilde{i}}\left|\frac{1}{T}\sum_{t=1}^{\tilde{t}}\delta_{it}\right|^2\right] \\ &\leq \frac{1}{N}\sum_{i=1}^{N}E\left[\max_{1\leq\tilde{t}\leq T}\left|\frac{1}{T}\sum_{t=1}^{\tilde{t}}\delta_{it}\right|^2\right]\leq\frac{M}{T}, \end{aligned}$$

having used convexity and Assumption 2. Thus, by virtue of Markov inequality, (36) holds if

$$\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{M}{\varepsilon^{2}NT^{2}}\frac{1}{\left[f_{3}\left(N\right)\right]^{2}\left[f_{4}\left(T\right)\right]^{2}}<\infty;$$

but this holds by virtue of the definitions of $f_3(N)$ and $f_4(T)$. Equation (36) entails that

$$\widehat{\bar{\lambda}}_{N}=\bar{\lambda}_{N}+O_{a.s.}\left[\frac{f_{3}\left(N\right)f_{4}\left(T\right)}{\sqrt{T}}\right],$$

and using this result in conjunction with (7) yields the desired result. QED.

Proof of Lemma 1. Consider, to begin with, two well-known results in matrix theory (we refer to Tao, 2012, for a comprehensive review) for details. First, the eigenvalue stability inequality stipulates that $\left|\lambda^{(i)}-\gamma^{(i)}\right|\leq\omega^{(1)}$ - usually this inequality is stated in terms of the operator norm of the perturbation matrix, which in our case is Σ_u. Second, note the dual Weyl inequality

$$\lambda^{(i+j-N)}\geq\gamma^{(i)}+\omega^{(j)},$$

which we use with $j=N$. Combining the two results we have

$$\gamma^{(i)}+\omega^{(N)}\leq\lambda^{(i)}\leq\gamma^{(i)}+\omega^{(1)}, \tag{37}$$

for $1\leq i\leq N$. Equation (5) is an immediate consequence of this result, and of Assumption 1*(i)*-*(ii)*. Consider (7). By (37)

$$\frac{1}{N}\sum_{i=1}^{N}\gamma^{(i)}+\omega^{(N)}\leq\bar{\lambda}_{N}\leq\frac{1}{N}\sum_{i=1}^{N}\gamma^{(i)}+\omega^{(1)}.$$

By Assumption 1*(i)*, $\frac{1}{N}\sum_{i=1}^{N}\gamma^{(i)}\leq M<\infty$; also, by part *(iii)* of the assumption we have

$$0<M''\leq\bar{\lambda}_{N}\leq M+\omega^{(1)},$$

which yields (7). QED.

Proof of Lemma 2. To being with, let us subtract (9) from (8). After some algebra we obtain

$$\left[\Sigma_{X}+\left(\hat{\Sigma}_{X}-\Sigma_{X}\right)\right]\left[x^{(i)}+\left(\hat{x}^{(i)}-x^{(i)}\right)\right]=\left[\lambda^{(i)}+\left(\hat{\lambda}^{(i)}-\lambda^{(i)}\right)\right]\left[x^{(i)}+\left(\hat{x}^{(i)}-x^{(i)}\right)\right];$$

multiplying both sides by $x^{(i)\prime}$, and recalling that $\left\|x^{(i)}\right\| = 1$ and $\Sigma_X x^{(i)} = \lambda^{(i)} x^{(i)}$ we obtain, after rearranging

$$x^{(i)\prime}\left(\hat{\Sigma}_X - \Sigma_X\right)\hat{x}^{(i)} = \left(\hat{\lambda}^{(i)} - \lambda^{(i)}\right)\left[1 + x^{(i)\prime}\left(\hat{x}^{(i)} - x^{(i)}\right)\right],$$

whence

$$\hat{\lambda}^{(i)} - \lambda^{(i)} = \frac{x^{(i)\prime}\left(\hat{\Sigma}_X - \Sigma_X\right)\hat{x}^{(i)}}{1 + x^{(i)\prime}\left(\hat{x}^{(i)} - x^{(i)}\right)}. \tag{38}$$

We start by showing that the denominator of (38) is $O_{a.s.}(1)$. Indeed, note that

$$\left|x^{(i)\prime}\left(\hat{x}^{(i)} - x^{(i)}\right)\right| \leq \left\|x^{(i)}\right\|\left\|\hat{x}^{(i)} - x^{(i)}\right\| \leq \left\|\hat{x}^{(i)}\right\| + \left\|x^{(i)}\right\| = 2, \tag{39}$$

by virtue of the Cauchy-Schwartz inequality and the triangle inequality. This is not the sharpest result, but it suffices for our purposes since it entails that $1 + x^{(i)\prime}\left(\hat{x}^{(i)} - x^{(i)}\right)$ is bounded almost surely.

Consider now the numerator. Let the j-th element of the vectors $x^{(i)}$ and $\hat{x}^{(i)}$ be defined as $x_j^{(i)}$ and $\hat{x}_j^{(i)}$ respectively, and let $\delta_{hjt} \equiv X_{ht}X_{jt} - \sigma_{hj}$. Then we can write

$$x^{(i)\prime}\left(\hat{\Sigma}_X - \Sigma_X\right)\hat{x}^{(i)} = \frac{1}{T}\sum_{h=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_h^{(i)}\hat{x}_j^{(i)}\delta_{hjt}.$$

We will show that

$$x^{(i)\prime}\left(\hat{\Sigma}_X - \Sigma_X\right)\hat{x}^{(i)} = O_{a.s.}\left[\frac{N}{\sqrt{T}} f_1^2(N) f_2(T)\right]. \tag{40}$$

As a preliminary result, we start by showing that

$$\sum_{N=1}^{\infty}\sum_{N=1}^{\infty}\sum_{T=1}^{\infty}\frac{1}{N^2 T} P\left[\max_{1\leq\tilde{h}\leq N, 1\leq\tilde{j}\leq N, 1\leq\tilde{t}\leq T}\left|\frac{1}{T}\sum_{h=1}^{\tilde{h}}\sum_{j=1}^{\tilde{j}}\sum_{t=1}^{\tilde{t}} x_h^{(i)}\hat{x}_j^{(i)}\delta_{hjt}\right| > \varepsilon\frac{N}{\sqrt{T}} f_1^2(N) f_2(T)\right] < \infty, \tag{41}$$

for some $\varepsilon > 0$. Equation (41) can be shown by noting that

$$\begin{aligned}
& \left| \frac{1}{T} \sum_{h=1}^{\tilde{h}} \sum_{j=1}^{\tilde{j}} \sum_{t=1}^{\tilde{t}} x_h^{(i)} \hat{x}_j^{(i)} \delta_{hjt} \right|^2 \\
\leq & \left[\sum_{h=1}^{\tilde{h}} \sum_{j=1}^{\tilde{j}} \left(x_h^{(i)} \right)^2 \left(\hat{x}_j^{(i)} \right)^2 \right] \left[\sum_{h=1}^{\tilde{h}} \sum_{j=1}^{\tilde{j}} \left(\frac{1}{T} \sum_{t=1}^{\tilde{t}} \delta_{hjt} \right)^2 \right] \\
\leq & \left[\sum_{h=1}^{N} \sum_{j=1}^{N} \left(x_h^{(i)} \right)^2 \left(\hat{x}_j^{(i)} \right)^2 \right] \left[\sum_{h=1}^{N} \sum_{j=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{\tilde{t}} \delta_{hjt} \right)^2 \right] \\
= & \sum_{h=1}^{N} \sum_{j=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{\tilde{t}} \delta_{hjt} \right)^2,
\end{aligned}$$

where the first inequality comes from the Cauchy-Schwartz inequality, and the last equality from $\left\| \hat{x}^{(i)} \right\| = \left\| x^{(i)} \right\| = 1$. This entails that

$$\begin{aligned}
& E \left[\max_{1 \leq \tilde{h} \leq N, 1 \leq \tilde{j} \leq N, 1 \leq \tilde{t} \leq T} \left| \frac{1}{T} \sum_{h=1}^{\tilde{h}} \sum_{j=1}^{\tilde{j}} \sum_{t=1}^{\tilde{t}} x_h^{(i)} \hat{x}_j^{(i)} \delta_{hjt} \right|^2 \right] \\
\leq & \sum_{h=1}^{N} \sum_{j=1}^{N} E \left[\max_{1 \leq \tilde{t} \leq T} \left(\frac{1}{T} \sum_{t=1}^{\tilde{t}} \delta_{hjt} \right)^2 \right] \leq M \frac{N^2}{T},
\end{aligned}$$

by virtue of Assumption 2. Thus, by Markov inequality, (41) holds if

$$\sum_{N=1}^{\infty} \sum_{N=1}^{\infty} \sum_{T=1}^{\infty} \frac{1}{N^2 T} \frac{T}{\varepsilon^2 N^2 \left[f_1^2 (N) f_2 (T) \right]^2} M \frac{N^2}{T} < \infty;$$

however, this follows from the definitions of $f_1(N)$ and $f_2(T)$. Hence

$$\lim \sup_{N,T \to \infty} \frac{\left| \sum_{h=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} x_h^{(i)} \hat{x}_j^{(i)} \delta_{hjt} \right|}{N \sqrt{T} f_1^2 (N) f_2 (T)} = 0 \text{ a.s.} \tag{42}$$

Equation (42) is shown, for the single index case, in Cai (2006). Showing that (41) implies (42) in the multi-index case is essentially the same, and we report the main passages hereafter. Note that, for every triple (N, N, T), there is a triple of positive integers (k_1, k_2, k_3) such that $2^{k_1} \leq N < 2^{k_1+1}$, $2^{k_2} \leq N < 2^{k_2+1}$, $2^{k_3} \leq T < 2^{k_3+1}$. Further, there is also a triple of real numbers defined over $[0,1)$, say (ρ_1, ρ_2, ρ_3), such that $N = 2^{k_1+\rho_1}$, etc... Define now the short-hand notation

$$L(k_1, k_2, k_3) \equiv \frac{\sqrt{2^{k_1+1}}\sqrt{2^{k_2+1}}}{\sqrt{2^{k_3+1}}} f_1\left(2^{k_1+\rho_1}\right) f_1\left(2^{k_2+\rho_2}\right) f_2\left(2^{k_3+\rho_3}\right),$$

$$S(k_1, k_2, k_3) \equiv \sum_{h=1}^{k_1}\sum_{j=1}^{k_2}\sum_{t=1}^{k_3} x_h^{(i)}\hat{x}_j^{(i)}\delta_{hjt}$$

and

$$P_{k_1,k_2,k_3} \equiv P\left[\max_{k_1,k_2,k_3} |S(k_1,k_2,k_3)| > \varepsilon L(k_1,k_2,k_3)\right],$$

where $\max_{k_1,k_2,k_3}$ is short for $\max_{1\leq k_1\leq 2^{k_1+\rho_1}, 1\leq k_2\leq 2^{k_2+\rho_2}, 1\leq k_3\leq 2^{k_3+\rho_3}}$. To being with, note first that (41) entails that

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}\frac{2^{k_1+1}2^{k_2+1}2^{k_3+1}}{(2^{k_1+1}-1)(2^{k_2+1}-1)(2^{k_3+1}-1)}P_{k_1,k_2,k_3} < \infty;$$

thus

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}P_{k_1,k_2,k_3} \leq \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}\frac{2^{k_1+1}2^{k_2+1}2^{k_3+1}}{(2^{k_1+1}-1)(2^{k_2+1}-1)(2^{k_3+1}-1)}P_{k_1,k_2,k_3} < \infty,$$

so that

$$\frac{\max_{k_1,k_2,k_3}|S(k_1,k_2,k_3)|}{L(k_1,k_2,k_3)} \to 0 \text{ a.s.}$$

Therefore

$$\begin{aligned}\frac{|S(N,N,T)|}{L(N,N,T)} &\leq \frac{\max_{k_1,k_2,k_3}|S(k_1,k_2,k_3)|}{L(k_1,k_2,k_3)}\frac{L(k_1,k_2,k_3)}{L(N,N,T)} \\ &\leq \sqrt{2}\frac{\max_{k_1,k_2,k_3}|S(k_1,k_2,k_3)|}{L(k_1,k_2,k_3)} \to 0 \text{ a.s.},\end{aligned}$$

so that

$$\lim\sup_{N,T\to\infty}\frac{|S(N,N,T)|}{L(N,N,T)} = 0 \text{ a.s.},$$

which proves (40). Equation (11) follows from combining (40) with (39). QED.

Proof of Theorem 1. Define the set

$$\Omega_{NT}^{(i)} \equiv \left\{\omega : \frac{\varphi^{(i)}}{\exp\left\{N^{\epsilon(1-\delta)}\right\}} > \varepsilon > 0\right\},$$

for some $0 < \epsilon < \frac{m_i}{\lambda_N}$ and any $\varepsilon > 0$. By Lemma 2, under H_0 with $\frac{N^{1-\delta}}{\sqrt{T}} \to 0$, $P\left[\lim_{N,T\to\infty} \Omega^{(i)}_{N,T}\right] = 1$. All the passages below are reported conditional on $\omega \in \Omega^{(i)}_{NT}$. For each u we have

$$\begin{aligned} \vartheta^{(i)}(u) &= \frac{2}{\sqrt{R}}\sum_{j=1}^{R}\left\{\zeta_j^{(i)}(u) - E^*\left[\zeta_j^{(i)}(u)\right]\right\} + \frac{2}{\sqrt{R}}\sum_{j=1}^{R}\left\{E^*\left[\zeta_j^{(i)}(u)\right] - \frac{1}{2}\right\} \\ &= I + II, \end{aligned} \tag{43}$$

with

$$E^*\left[\zeta_j^{(i)}(u)\right] = \frac{1}{2} + \frac{1}{\sqrt{2\pi\varphi^{(i)}}}\int_0^u \exp\left[-\frac{1}{2}\frac{t^2}{\varphi^{(i)}}\right]dt, \tag{44}$$

where we consider the case of $u > 0$ without loss of generality. Consider first II in (43); based on (44), we have

$$II = \frac{2\sqrt{R}}{\sqrt{2\pi}}\int_0^{u/\sqrt{\varphi^{(i)}}} \exp\left[-\frac{1}{2}s^2\right]ds = \frac{\sqrt{2R}}{\sqrt{\pi}}\frac{u}{\sqrt{\varphi^{(i)}}}\left[1 - \frac{1}{6}b^3\right],$$

for some $b \in \left(0, \frac{u}{\sqrt{\varphi^{(i)}}}\right)$. Note that this expression is bounded from above by $\sqrt{\frac{2\bar{u}^2}{\pi}}\sqrt{\frac{R}{\varphi^{(i)}}}$, where $\bar{u}$ is the boundary value of U. Under the null that $\lambda^{(i)} = m_i N$, when $\frac{N^{1-\delta}}{\sqrt{T}} \to 0$, it holds that $\varphi^{(i)} = \exp\left\{2\frac{m_i}{\lambda_N}N^{1-\delta}\right\}\exp\left\{o_{a.s.}(1)\right\}$; hence, II is $o_p(1)$ uniformly in u. Turning to I in (43), $\bar{\zeta}_j^{(i)}(u) \equiv \zeta_j^{(i)}(u) - E^*\left[\zeta_j^{(i)}(u)\right]$ is an *i.i.d.* sequence with mean zero and

$$V^*\left[\frac{2}{\sqrt{R}}\sum_{j=1}^{R}\bar{\zeta}_j^{(i)}(u)\right] = 4E^*\left[\frac{1}{R}\sum_{j=1}^{R}\sum_{k=1}^{R}\bar{\zeta}_j^{(i)}(u)\,\bar{\zeta}_k^{(i)}(u)\right] = 4\frac{1}{R}\sum_{j=1}^{R}E^*\left[\left(\bar{\zeta}_j^{(i)}(u)\right)^2\right] = 1 + o_p(1),$$

where the second equality comes from the fact that $\bar{\zeta}_j^{(i)}(u)$ is generated independently across j, and the last equality comes from (44) and the passages thereafter. This holds uniformly in u by the same passages as above. Thus, a CLT can be applied to I, so that, as $(N,R) \to \infty$, $I \xrightarrow{d^*} N(0,1)$. Putting everything together, as $(N,R) \to \infty$ with (20), $\vartheta^{(i)}(u) \xrightarrow{d^*} N(0,1)$ uniformly in u. This entails that $\lim_{N,R\to\infty}\Theta^{(i)} = \lim_{N,R\to\infty}\int_{\underline{u}}^{\overline{u}}\left[\vartheta^{(i)}(u)\right]^2\varphi(u)\,du \xrightarrow{d^*} \int_{\underline{u}}^{\overline{u}}\chi_1^2\varphi(u)\,du = \chi_1^2$. QED

Proof of Theorem 2. Define

$$\Omega^{+(i)}_{NT} \equiv \left\{\omega : \varphi^{(i)} \le M < \infty\right\},$$

such that under H_A we have $P\left[\lim_{N\to\infty}\Omega^{+(i)}_{NT}\right] = 1$. All the passages below are reported conditional on $\omega \in \Omega^{+(i)}_{NT}$. Consider (43). Term I still satisfies a CLT by construction, so that, under

H_A, $I \overset{d^*}{\to} N(0,1)$. As far as II in (43) is concerned, recall that

$$II = \frac{2\sqrt{R}}{\sqrt{2\pi}} \int_0^{u/\sqrt{\varphi^{(i)}}} \exp\left[-\frac{1}{2}s^2\right] ds = \frac{\sqrt{2R}}{\sqrt{\pi}} \frac{u}{\sqrt{\varphi^{(i)}}} \left[1 - \frac{1}{6}b^3\right];$$

thus, $\vartheta^{(i)}(u)$ has a non-centrality parameter proportional to $\sqrt{R}$. Hence, $\Theta^{(i)}$ has a noncentrality parameter that diverges to positive infinity as long as $R \to \infty$, which gives the desired result. QED

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica, 81, 1203-1227.

Alessi, L., M. Barigozzi, and M. Capasso, 2010. Improved penalization for determining the number of factors in approximate factor models. Statistics and Probability Letters, 80, 1806-1813.

Amengual, D., 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, 91-96.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica, 70, 191-221.

Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business & Economic Statistics, 25, 52-60.

Bandi, F., Corradi, V., 2014. Nonparametric nonstationarity tests. Econometric Theory, 30, 127-149.

Cai, G., 2006. Chover-type laws of the iterated logarithm for weighted sums of $\tilde{\rho}$-mixing sequences, Journal of Applied Mathematics and Stochastic Analysis 2006, Article ID 65023.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 51, 1281-1304.

Corradi, V., Swanson, N.R., 2002. A consistent test for nonlinear out of sample predictive accuracy. Journal of Econometrics, 110, 353–381.

Corradi, V., Swanson, N.R., 2006. The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. Journal of Econometrics, 132, 195-229.

Davidson, J., 2002. Stochastic limit theory. Oxford: Oxford University Press.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Review of Economics and Statistics, 82, 540-554.

Hallin, M., Liska, R., 2007. Determining the number of factors in the generalized dynamic factor model. Journal of the American Statistical Association, 102, 603-617.

Merlevede, F., Peligrad, M., 2013. Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. Annals of Probability, 41, 914-960.

Onatski, A., 2009. A formal statistical test for the number of factors in the approximate factor models. Econometrica, 77, 1447-1479.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economic and Statistics, 92, 1004-1016.

Pearson, E.S., 1950. On questions raised by the combination of tests based on discontinuous distributions. Biometrika, 37, 383-398.

Rio, E., 1995. A maximal inequality and dependent Marcinkiewicz-Zygmund strong laws. Annals of Probability, 23, 918-937.

Rosenthal, H.P., 1970. On the subspaces of L_p $(p > 2)$ spanned by sequences of independent random variables. Israel Journal of Mathematics, 8, 273–303.

Tao, T., 2012. *Topics in random matrix theory.* American Mathematical Society.

T \ N	25	50	75	100	250
50	0.062	0.062	0.060	0.049	0.058
100	0.056	0.060	0.066	0.043	0.043
250	0.051	0.065	0.060	0.047	0.044
500	0.050	0.056	0.065	0.041	0.056
1000	0.064	0.041	0.060	0.055	0.055

Table 1a. Empirical rejection frequencies for tests for $H_0 : \lambda^{(2)} \to \infty$ as $(N, T) \to \infty$ when $k = 2$. Case of no serial or cross sectional dependence.

T \ N	25	50	75	100	250
50	0.054	0.062	0.060	0.049	0.058
100	0.056	0.062	0.066	0.043	0.043
250	0.050	0.066	0.060	0.047	0.044
500	0.050	0.056	0.065	0.041	0.056
1000	0.065	0.041	0.060	0.055	0.055

Table 1b. Empirical rejection frequencies for tests for $H_0 : \lambda^{(2)} \to \infty$ as $(N, T) \to \infty$ when $k = 2$. Case of serial dependence.

T \ N	25	50	75	100	250
50	0.052	0.063	0.057	0.049	0.058
100	0.053	0.059	0.066	0.043	0.043
250	0.052	0.065	0.060	0.047	0.044
500	0.050	0.056	0.065	0.041	0.056
1000	0.065	0.041	0.060	0.055	0.055

Table 1c. Empirical rejection frequencies for tests for $H_0 : \lambda^{(2)} \to \infty$ as $(N, T) \to \infty$ when $k = 2$. Case of cross sectional dependence.

T \ N	25	50	75	100	250
50	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000
250	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000

Table 2a. Empirical rejection frequencies for tests for $H_0 : \lambda^{(2)} \to \infty$ as $(N, T) \to \infty$ when $k = 1$. Case of no serial or cross sectional dependence.

T \ N	25	50	75	100	250
50	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000
250	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000

Table 2b. Empirical rejection frequencies for tests for $H_0 : \lambda^{(2)} \to \infty$ as $(N, T) \to \infty$ when $k = 1$. Case of serial dependence.

T \ N	25	50	75	100	250
50	0.460	0.958	0.999	1.000	1.000
100	0.479	0.953	1.000	1.000	1.000
250	0.459	0.996	0.999	1.000	1.000
500	0.327	0.982	1.000	1.000	1.000
1000	0.724	0.896	1.000	1.000	1.000

Table 2c. Empirical rejection frequencies for tests for $H_0 : \lambda^{(2)} \to \infty$ as $(N, T) \to \infty$ when $k = 1$. Case of cross sectional dependence.

	N 25	50	75	100	250
T					
50	2.951	.	.	.	.
.	[0.028]	.	.	.	.
100	2.964	2.955	2.786	.	.
.	[0.021]	[0.035]	[0.203]	.	.
250	2.891	2.969	2.964	2.978	.
.	[0.101]	[0.014]	[0.017]	[0.012]	.
500	2.960	2.966	2.966	2.969	2.971
.	[0.035]	[0.018]	[0.017]	[0.016]	[0.013]
1000	2.986	2.970	2.966	2.959	2.977
.	[0.009]	[0.013]	[0.017]	[0.022]	[0.013]

Table 3a. Average estimate of the number of factors and (in square brackets) percentage of wrong estimates - case of no serial or cross sectional dependence and $k = 3$.

	N 25	50	75	100	250
T					
50	2.943	.	.	.	.
.	[0.058]	.	.	.	.
100	2.976	2.958	2.764	.	.
.	[0.017]	[0.031]	[0.203]	.	.
250	2.986	2.965	2.964	2.969	.
.	[0.010]	[0.016]	[0.017]	[0.016]	.
500	2.953	2.966	2.966	2.969	2.971
.	[0.033]	[0.018]	[0.017]	[0.017]	[0.013]
1000	2.973	2.970	2.966	2.959	2.977
.	[0.017]	[0.013]	[0.017]	[0.022]	[0.013]

Table 3b. Average estimate of the number of factors and (in square brackets) percentage of wrong estimates - case of serial dependence and $k = 3$.

	N	50	75	100	250
T					
50	.	.	.	.	.
.	.	.	.	.	.
100	.	2.971	2.786	.	.
.	.	[0.018]	[0.187]	.	.
250	.	2.977	2.967	2.945	.
.	.	[0.030]	[0.014]	[0.026]	.
500	.	3.002	2.966	2.968	2.971
.	.	[0.042]	[0.017]	[0.017]	[0.012]
1000	.	2.986	2.966	2.959	2.977
.	.	[0.030]	[0.018]	[0.020]	[0.013]

Table 3c. Average estimate of the number of factors and (in square brackets) percentage of wrong estimates - case of cross sectional dependence and $k = 3$.