



Determining the number of factors when the number of factors can increase with sample size



Hongjun Li^a, Qi Li^{b,a}, Yutang Shi^{c,d,*}

^a International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, PR China

^b Department of Economics, Texas A&M University, USA

^c School of Finance, Nankai University, Tianjin, 300350, PR China

^d Wells Fargo & Company, USA

ARTICLE INFO

Article history:

Received 17 December 2013

Received in revised form

31 May 2016

Accepted 5 June 2016

Available online 14 November 2016

JEL classification:

C2

C3

C5

G1

Keywords:

Principal components

Factor analysis

Increasing number of factors

Information criteria

ABSTRACT

Correctly specifying the number of factors (r) is a fundamental issue for the application of factor models. In this paper we develop an econometric method to estimate the number of factors in factor models of large dimensions where the number of factors is allowed to increase as the two dimensions, cross-section size (N) and time period (T) increase. Using similar information criteria as proposed by Bai and Ng (2002), we show that the number of factors can be consistently estimated using the criteria. We propose a new procedure that avoids over estimating the number of factors while allowing for one to search for possible number of factors over a wide range of positive integers so that it also avoids underestimation of the number of factors. We conduct Monte-Carlo simulation to investigate the finite sample properties of the proposed approach.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Factor models have been widely used in economic analyses such as forecasting economic variables, estimating variance–covariance matrices with high dimensional data, and estimating average treatment effects, among others. In practice a few common factors may capture the variations of a large number of economic variables. In the finance literature, the arbitrage pricing theory (APT) of Ross (1976) assumes that a small number of factors can be used to explain a large number of asset returns. Stock and Watson (1998) and Stock and Watson (1999) consider forecasting inflation with diffusion indices (“factors”) constructed from a large number of macroeconomic series. Gregory and Head (1999) and Forni et al. (2000) find that cross country variations have common

components. Fan et al. (2011, 2013) use factor models to estimate high dimensional variance–covariance matrices. Factor models can also be used to evaluate the impacts of various policies (e.g., Hsiao et al. (2012)). By assuming that the cross-sectional correlations for all the units are attributed to the presence of some (unobserved) common factors, Hsiao et al. (2012), Ching et al. (2012), Bai et al. (2014), Ouyang and Peng (2015) and Li and Bell (2017) use panel data methods to construct the counterfactuals and to measure average treatment effects of some policy interventions based on factor models.

A fundamental issue of factor models is the correct specification of the number of factors, r . When the number of factors is fixed, Bai and Ng (2002), Onatski (2009), Ahn and Horenstein (2013), among others, have developed various approaches to consistently estimate the number of factors. But many empirical findings suggest that the number of factors may increase as the dimensions of the data N increases, or T increases. For many empirical analyses, the estimated number of factors ranges from one to more than ten, see Ludvigson and Ng (2009), Giannone et al. (2005) and Forni and Gambetti (2010). This suggests that the

* Corresponding author at: School of Finance, Nankai University, Tianjin, 300350, PR China.

E-mail addresses: econlee@163.com (H. Li), qi-li@tamu.edu (Q. Li), shij_2001@hotmail.com (Y. Shi).

number of factors may depend on sample sizes. One reason that the number of factors may increase with sample size is structural break, new factors may emerge after economic environments change. Using Bai and Ng’s information criteria, Ludvigson and Ng (2007) find that the factor structure of their financial dataset comprising of 172 ($N = 172$) series quarterly financial indicators spanning the first quarter of 1960 through the fourth quarter of 2002 ($T = 172$) can be well described by 8 ($r = 8$) common factors. Jurado et al. (2015) update monthly version of the 147 financial time series used in Ludvigson and Ng (2007) and combine them with an updated version of 132 monthly macroeconomic series used in Ludvigson and Ng (2010). They find that 12 ($r = 12$) common factors can capture the variations of this new dataset with 279 series ($N = 279$) spanning the period 1959:01–2011:12 ($T = 636$). Hence, finding in Jurado et al. (2015) supports the argument that the number of factors may increase as sample size increases.

Assuming that the number of factors r is fixed, there are many papers in the literature analyzing the problem of determining the number of factors. Some of them not only fix the number of factors, but also impose restrictions on the dimensions N and T , such as Lewbel (1991), Donald (1997), Cragg and Donald (1997), Connor and Korajczyk (1993), Forni and Reichlin (1998) and Stock and Watson (1998). Imposing no restriction on the relation between N and T except that both N and T are assumed to be large, Bai and Ng (2002) treat the determination of the number of factors as a model selection problem, they propose some criteria and show that the number of factors can be consistently estimated by minimizing the proposed criteria. To improve the finite sample performances of Bai and Ng’s criteria, Hallin and Liska (2007), and Alessi et al. (2010) propose some modifications to Bai and Ng’s procedure by introducing a tuning multiplicative constant in the penalty objective function. Onatski (2009) develops a test of the null of k_0 factors against the alternative that the number of factors r satisfies $k_0 < r \leq k_1$ for some finite positive integer k_1 . Onatski also describes the asymptotic distribution of the test statistic with critical values tabulated. Onatski (2010) suggests to determine the number of factors from empirical distribution of eigenvalues of sample covariance matrix. Ahn and Horenstein (2013) exploit the fact that the r largest eigenvalues of the variance matrix of N response variables grow unboundedly as N increases, while the other eigenvalues remain bounded to estimate the number of factors. All of the above mentioned works consider the case of a fixed number of factors (r is fixed). The main difference between our paper and the existing work is that we consider the problem of determining the number of factors in a factor model where the number of factors is allowed to increase as N or T increases.

Specifically, this paper is designed to provide an approach which enables one to estimate the number of factors consistently when the number of factors is allowed to increase as $N, T \rightarrow \infty$. We extend the method of Bai and Ng (2002) to penalize the number of factors with a penalty function which is determined by the sample sizes, N and T , as well as the maximum possible number of factors allowed in the estimation. As the factors are unobserved, the estimation procedure takes two steps. First, assuming the number of factors to be an arbitrary number $1 \leq k \leq k_{max}$, we estimate the factors (\hat{F}^k) using the principal components method, where $k_{max} = k_{max,N,T}$ is the maximum number for possible number of factors, which is assumed to be greater than or equal to the true number of factors, whose value is determined by N and T and it increases as N, T increases. Second, we select the number of factors \hat{k} by minimizing a criterion modified from Bai and Ng (2002), which is a function of k and the estimated factors (\hat{F}^k). This criterion depends on the usual trade-off between good fit and parsimony. We show that this method produces a consistent estimator of the number of factors r . However, simulation results

show that the selected number of factors \hat{k} can be sensitive to the choice of k_{max} and it tends to choose a \hat{k} that is larger than r when k_{max} is large. We propose using a new (‘mode’ based) selection procedure to overcome this problem so that the selected \hat{k} is not sensitive to different k_{max} values used in practice.

The rest of this paper is organized as follows. Section 2 sets up the model and presents the assumptions associated with the model. Section 3 presents the estimating procedures and the theoretical properties of the proposed estimators. Section 4 reports simulation experiments to examine the finite sample performances of our proposed method when r increases with N or T . Concluding remarks are given in Section 5. All the proofs are given in the Appendix.

2. Factor models

We consider the problem of determining the number of factors (r) in a static approximate factor model, allowing $r = r_{N,T} \rightarrow \infty$, as $N \rightarrow \infty$, or $T \rightarrow \infty$, or both $N, T \rightarrow \infty$, but with a slower rate than $\min\{N, T\}$, i.e., $\max\{r/N, r/T\} \rightarrow 0$, as $N, T \rightarrow \infty$.

Let X_{it} denote the response variable for unit i at time t , for $i = 1, \dots, N$, and $t = 1, \dots, T$. Our model is of the following form

$$X_{it} = \frac{1}{\sqrt{r}} \lambda_i^0 F_t^0 + e_{it}, \tag{1}$$

where F_t^0 is an $r \times 1$ vector of common factors, λ_i^0 is the $r \times 1$ vector of factor loadings, and e_{it} is the idiosyncratic error of the response variable X_{it} . The factors, factor loadings and idiosyncratic errors are not observed. Without loss of generality, we can assume that $E(X_{it}) = 0$. If this is not the case, we can de-mean the data first.

Note that at the right-hand-side of our model (1), we divide $\lambda_i^0 F_t^0$ by \sqrt{r} . This is because we allow for r to diverge when $N, T \rightarrow \infty$. If we do not divide $\lambda_i^0 F_t^0$ by \sqrt{r} , then the variance of the systematic part, $\lambda_i^0 F_t^0$, is proportional to r and the variance of idiosyncratic error e_{it} is finite, the variance of noise part over the variance of information part will go to zero, or equivalently, the goodness-of-fit R^2 will converge to one. By dividing $\lambda_i^0 F_t^0$ by \sqrt{r} , we have $Var(r^{-1/2} \lambda_i^0 F_t^0) = O(1)$ and we can obtain a reasonable goodness-of-fit that is not too close to one.

Let $\text{tr}(A)$ denote the trace of a square matrix A . The norm of a matrix A is defined as $\|A\| = [\text{tr}(A'A)]^{1/2}$. We use M_1 to denote a generic positive constant and use \mathcal{N} to denote the set of natural numbers. We make the main assumptions as follows:

Assumption A (Factors and Loadings).

1. For all $t, r^{-2} E \|F_t^0\|^4 < M_1$;
2. There exists a $r \times r$ positive definite matrix Σ_F such that $\|T^{-1} \sum_{t=1}^T F_t^0 F_t^{0'} - \Sigma_F\| \xrightarrow{P} 0$ as $T \rightarrow \infty$;
3. $\max_{1 \leq i \leq N} r^{-2} E \|\lambda_i^0\|^4 \leq M_1 < \infty$;
4. Let Λ^0 be the $N \times r$ factor loading matrix with its i th row given by λ_i^0 . Then there exists a $r \times r$ positive definite matrix D such that $\|N^{-1} \Lambda^{0'} \Lambda^0 - D\| \xrightarrow{P} 0$ as $N \rightarrow \infty$;
5. Let λ_{il}^0 and F_{tl}^0 be the l th components ($l = 1, \dots, r$) of λ_i^0 and F_t^0 , respectively. Then for all (i, t) ,

$$E \{ [r^{-1/2} \sum_{l=1}^r E(\lambda_{il}^0 F_{tl}^0)]^4 \} \leq M_1.$$

Assumption B (Idiosyncratic Components). As $N, T \rightarrow \infty$ and $0 < \lim_{N,T \rightarrow \infty} \frac{T}{N} < \infty$,

1. For all i and t , $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M_1$;
2. $E(N^{-1}e'_s e_t) = E(N^{-1} \sum_{i=1}^N e_{is} e_{it}) = \gamma_N(s, t)$, $|\gamma_N(s, s)| \leq M_1$ for all s , and that

$$T^{-1} \sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s, t)| \leq M_1;$$

3. $E(e_{it} e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t ; furthermore, $N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}| \leq M_1$;
4. $E(e_{it} e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\tau_{ij,ts}| \leq M_1$;
5. for every (t, s) , $E|N^{-1/2} \sum_{i=1}^N [e_{is} e_{it} - E(e_{is} e_{it})]|^4 \leq M_1$;
6. We assume that there exist a $T \times T$ matrix L , a $N \times N$ matrix R , and a $T \times N$ matrix ε such that

$$e = L\varepsilon R$$

where L ($T \times T$) and R ($N \times N$) are arbitrary non-random positive definite matrices, and $\varepsilon = (\varepsilon_{ti})$ is a $T \times N$ matrix consisting of independent elements with uniformly bounded 7th moment and $E(\varepsilon_{it}) = 0$.

Assumption C. 1. Weak Dependence Between Factors and Idiosyncratic Components:

$$E \left(\frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t^0 e_{it} \right\|^2 \right) \leq M_1;$$

2. Weak Dependence Between Factor Loadings and Idiosyncratic Components:

$$E \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{Nr}} \sum_{i=1}^N e_{it} \lambda_i^0 \right\|^2 \right) \leq M_1.$$

Conditions in Assumption A are modified from Assumptions A–B in Bai and Ng (2002) by taking care of the fact that $r \rightarrow \infty$ as $N, T \rightarrow \infty$. It is easy to see that Assumption A1 holds true if $E[(F_{it}^0)^4] = O(1)$ for all $t = 1, \dots, T$ and for all $l = 1, \dots, r$. A2 imposes a restriction on the rate of r . For example with $\Sigma_F = T^{-1} \sum_{t=1}^T E(F_t^0 F_t^{0'})$, it can be easily shown that A2 holds true if $r = o(T^{1/2})$ and $\sum_{t=1}^T \sum_{s \neq t}^T \text{Cov}(F_{tm}^0 F_{tm}^0, F_{sl}^0 F_{sl}^0) = O(T)$ for all $l, m \in \{1, \dots, r\}$. This is because $E[\|T^{-1} \sum_{t=1}^T F_t^0 F_t^{0'} - \Sigma_F\|^2] = T^{-2} \sum_{t=1}^T \sum_{s=1}^T \sum_{l=1}^r \sum_{m=1}^r E\{[F_{tl}^0 F_{tm}^0 - E(F_{tl}^0 F_{tm}^0)][F_{sl}^0 F_{sm}^0 - E(F_{sl}^0 F_{sm}^0)]\} = O(r^2/T)$ if for all $l, m = 1, \dots, r$, it holds that $\sum_{t=1}^T \sum_{s \neq t}^T E\{[F_{tl}^0 F_{tm}^0 - E(F_{tl}^0 F_{tm}^0)][F_{sl}^0 F_{sm}^0 - E(F_{sl}^0 F_{sm}^0)]\} = O(T)$. Similarly, A3 holds true if $E[(\lambda_{it}^0)^4] = O(1)$ for all $i = 1, \dots, N$ and for all $l = 1, \dots, r$. A4 is similar to A2, it holds true with $D = N^{-1} \sum_{i=1}^N E(\lambda_i^0 \lambda_i^0)$, if $r = o(N^{1/2})$ and $\sum_{i=1}^N \sum_{j \neq i}^T \text{Cov}(\lambda_{il}^0 \lambda_{jm}^0, \lambda_{jl}^0 \lambda_{im}^0) = O(N)$ for all $l, m \in \{1, \dots, r\}$. A5 requires that $\lambda_{it}^0 F_{it}^0$ is a weakly dependent process in l because we allow for $r \rightarrow \infty$.

Conditions in Assumption B are basically the same as Assumption C in Bai and Ng (2002) because the idiosyncratic error e_{it} is unrelated to r whether r is finite or is allowed to diverge to infinity with the sample size. In particular, B5 is similar to A5 in that it assumes that, for all (t, s) , $e_{it} e_{is}$ is a weakly dependent process in i . Assumption B6 puts a structure on the idiosyncratic components. This structure allows heteroscedasticity in both the time and cross-section dimensions, and also limited autocorrelation and cross-sectional correlation in the components.

Finally, Assumption C is similar to assumption D in Bai and Ng (2002) except that we modified it by dividing the quantity by \sqrt{r} as r is allowed to diverge as N and T tend to infinity. They allow for limited time-series and cross-section dependence in idiosyncratic component and also weak dependence between factors (factor loadings) and idiosyncratic errors.

3. Estimating the common factors and the number of factors

Following Bai and Ng (2002), we estimate the common factor in a large panel by the principal components method. For $k \in \{1, \dots, k_{max}\}$, where k_{max} is allowed to increase at a slower speed than $\min\{N, T\}$ such that $k_{max} = o(\min\{N^{1/3}, T\})$. Let λ_i^k and F_t^k denote $k \times 1$ vectors of the loadings and factors with the allowance of k factors in the estimation. The method of principal components minimizes

$$V(k) = \min_{\Lambda^k, F^k} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(X_{it} - \frac{1}{\sqrt{k}} \lambda_i^k F_t^k \right)^2 \tag{2}$$

over $1 \leq k \leq k_{max}$, subject to the normalization of either $\Lambda^k \Lambda^k / N = I_k$ or $F^k F^k / T = I_k$, where Λ^k and F^k are the $N \times k$ and $T \times k$ factor loading and factor matrices, respectively.

Let $ev_{(i)}(A)$ denote the i th largest eigenvalue of matrix A , and $EV_{(i)}(A)$ is the eigenvector corresponding to the eigenvalue $ev_{(i)}(A)$ of the matrix A . If one concentrates out Λ^k and uses the normalization that $F^k F^k / T = I_k$, the estimated factor matrix is $\tilde{F}^k = \sqrt{T}(EV_1(XX'), \dots, EV_k(XX'))$. Given \tilde{F}^k , $\tilde{\lambda}^k = \sqrt{k}(\tilde{F}^k \tilde{F}^k)^{-1} \tilde{F}^k X = \sqrt{k} \tilde{F}^k X / T$ is the corresponding matrix of factor loadings. On the other hand, if one concentrates out F^k and uses the normalization that $\Lambda^k \Lambda^k / N = I_k$, the solution to the above problem is given by $(\tilde{F}^k, \tilde{\lambda}^k)$, where $\tilde{\lambda}^k = \sqrt{N}(EV_{(1)}(X'X), \dots, EV_{(k)}(X'X))$. The normalization that $\Lambda^k \Lambda^k / N = I_k$ implies $\tilde{F}^k = \sqrt{k} X \tilde{\lambda}^k / N$.

Define $\hat{F}^k = \tilde{F}^k (\tilde{F}^k \tilde{F}^k / T)^{1/2}$, a rescaled estimator of the factors. This rescaled estimator has the asymptotic properties summarized in the following theorem.

Proposition 1. Under Assumptions A–C, for any $1 \leq k \leq k_{max} = o(\min\{N^{1/3}, T\})$ there exists a $(r \times k)$ matrix H^k with rank = $\min\{k, r\}$ such that

$$\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_t^k - H^k F_t^0 \right\|^2 = O_p \left(\max \left\{ \frac{k^3 r}{N}, \frac{k^3}{T} \right\} \right). \tag{3}$$

Similar to the results of Bai and Ng (2002), Proposition 1 suggests that the time average of the squared deviations between the estimated factors \hat{F}^k and those that lie in the true factor space, $H^k F_t^0$, will vanish as $N, T \rightarrow \infty$. However, the convergence rate depends on not only the panel structure N and T , but also the factor structure r and k .

Given the results of Proposition 1, we can now analyze the problem of determining the number of factors. Let $V(k, F^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \frac{1}{\sqrt{k}} \lambda_i^k F_t^k)^2$ be the sum of squared residuals (divided by NT), where the residuals are from regression models of regressing X_i on the k factors for all $i = 1, \dots, N$, and $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})'$ is a $T \times 1$ vector of time-series observations for the i th cross-section unit. The selecting criterion modified from those suggested by Bai and Ng (2002) has the form

$$PC(k) = V(k, \hat{F}^k) + kg(N, T), \tag{4}$$

where $g(N, T)$ is the penalty factor satisfying two conditions: (i) $k_{max} \cdot g(N, T) \rightarrow 0$ as $N, T \rightarrow \infty$, (ii) $C_{N,T,k_{max}}^{-1} g(N, T) \rightarrow \infty$ as $N, T \rightarrow \infty$, where $C_{N,T,k_{max}} = O_p \left(\max \left\{ \frac{k_{max}^3}{\sqrt{N}}, \frac{k_{max}^{5/2}}{\sqrt{T}} \right\} \right)$.

As $V(k, \hat{F}^k)$ is decreasing in k , the criterion above penalizes k with a penalty factor $kg(N, T)$ to select the estimator \hat{k} such that asymptotically under and overparameterized models will not be chosen. Theorem 1 formally establishes this result.

Theorem 1. Let $1 \leq r \leq k_{max} = o(\min\{N^{1/17}, T^{1/16}\})$ and $\hat{k} = \operatorname{argmin}_{1 \leq k \leq k_{max}} PC(k)$. Suppose that Assumptions A–C hold, and that (i) $k_{max} \cdot g(N, T) \rightarrow 0$, (ii) $C_{N,T,k_{max}}^{-1} \cdot g(N, T) \rightarrow \infty$ as $N, T \rightarrow \infty$. Then

$$\lim_{N, T \rightarrow \infty} \operatorname{Prob}[\hat{k} = r] = 1. \tag{5}$$

A formal proof of Theorem 1 is provided in the Appendix. Conditions (i) and (ii) together define the type of penalty factor that should vanish at an appropriate rate. They are sufficient conditions for estimation consistency so that they may not always be required for consistently estimating the number of factors.

Remark 1. Since we often need to divide some quantities by r , we rule out the case that $r = 0$. Allowing for $r = 0$ in our framework will complicate the regularity conditions, notations and proofs. Therefore, we did not consider the case that $r = 0$ in our paper. The $r = 0$ case is covered in Bai and Ng’s (2002). Their procedure can be used to select the number of factors even when the true number of factors is 0. We also conducted some simulations which show that both Bai and Ng’s (2002) original method and the modified method proposed in our paper work well when $r = 0$.

Note that the condition imposed in k_{max} is asymmetric in (N, T) . This result is induced by Proposition 1. The details can be found in the proof of Proposition 1 given in the Appendix. As a referee correctly points out, if in the proof of Theorem 1, instead of using the result of Proposition 1 that $T^{-1} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 = O_p\left(\max\left\{\frac{k^3 r}{N}, \frac{k^3}{T}\right\}\right)$, one may use $N^{-1} \sum_{i=1}^N \|\hat{\lambda}_i^k - \tilde{H}^k \lambda_i^0\|^2 = O_p\left(\max\left\{\frac{k^3}{N}, \frac{k^3 r}{T}\right\}\right)$,¹ where \tilde{H}^k is a $r \times k$ matrix with $\operatorname{rank}(\tilde{H}^k) = \min\{r, k\}$. Then the condition that $1 \leq r \leq k_{max} = o(\min\{N^{1/17}, T^{1/16}\})$ in Theorem 1 will be replaced by $1 \leq r \leq k_{max} = o(\min\{N^{1/16}, T^{1/17}\})$. The result is still asymmetric in N and T , but the roles of N and T are exchanged.

In fact, it is possible to obtain a symmetric result (of k_{max} in N and T) under some stronger regularity conditions, i.e., one can obtain a symmetric condition for k_{max} as $1 \leq r \leq k_{max} = o(\min\{N^{1/16}, T^{1/16}\})$ in Theorem 1 under some stronger assumptions. We state this result in the following proposition.

Proposition 2. Under the same conditions as in Proposition 1 except that we strength some conditions as follows: (i) λ_{il} is non-random with $\lambda_{il} \leq \bar{\lambda} < \infty$ for all $i = 1, \dots, N$ and $l = 1, \dots, r$; (ii) $E(e_{it} e_{jt}) = 0$ for all $t \in \{1, \dots, T\}$ and for all $j \neq i, i, j \in \{1, \dots, N\}$, $E(F_{it}^0 F_{tm}^0) = 0$ for all $t \in \{1, \dots, T\}$ and for all $m \neq l, l, m \in \{1, \dots, r\}$; (iii) e_{it} and F_s^0 are independent with each other for all i, t and s . Then

$$\frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 = O_p\left(k^3 \left(\frac{1}{N} + \frac{1}{T}\right)\right). \tag{6}$$

The proof of Proposition 2 is given in the appendix. Under Proposition 2, the condition $1 \leq r \leq k_{max} = o(\min\{N^{1/17}, T^{1/16}\})$ in Theorem 1 can be replaced by $1 \leq r \leq k_{max} = o(\min\{N^{1/16}, T^{1/16}\})$. That is, we obtain a condition on k_{max} that is symmetric in N and T .

¹ This result can be proved similar to the proof of Proposition 1, its proof is available from the authors upon request.

Remark 2. The zero correlation assumption on F_{it} used in Proposition 2 is quite strong. However, it can be replaced by some weakly dependence assumptions such as ρ -mixing or β -mixing processes with mixing coefficients decay to zero at certain rates. But this will make the presentation (regarding regularity conditions) as well as the proofs of Proposition 2 much longer. Therefore, we will not pursue a proof of Proposition 2 under weak regularity conditions in this paper.

Similar to Bai and Ng (2002) we have the following Corollary.

Corollary 1. Under the Assumptions of Theorem 1, if one replaces $PC(k)$ in Theorem 1 by the class of criteria defined by

$$IC(k) = \ln(V(k, \hat{F}^k)) + kg(N, T),$$

then the conclusion of Theorem 1 holds true.

Corollary 3.1 states that the class of criteria $IC(k)$ can also be used to consistently estimate the number of factors in factor models where the number of factors possibly increases with the sample size.

Let $\hat{\sigma}^2$ be a consistent estimate of $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E(e_{it})^2$. Bai and Ng (2002) generalize the C_p criterion of Mallows (1973) and suggest three PC_p criteria as follows:

$$\begin{aligned} PC_{p1}(k) &= V(k, \hat{F}^k) + k \cdot \hat{\sigma}^2 \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right), \\ PC_{p2}(k) &= V(k, \hat{F}^k) + k \cdot \hat{\sigma}^2 \left(\frac{N+T}{NT}\right) \ln(\min\{N, T\}), \\ PC_{p3}(k) &= V(k, \hat{F}^k) + k \cdot \hat{\sigma}^2 \left(\frac{\ln(\min\{N, T\})}{\min\{N, T\}}\right). \end{aligned} \tag{7}$$

It is easy to check that these criteria satisfy the two conditions for the penalty factor in Theorem 1 if $k_{max} = o\left(\left[\ln\left(\frac{NT}{N+T}\right)\right]^{1/6}\right)$. The three criteria have different finite-sample properties while they are asymptotically equivalent. In applications, Bai and Ng (2002) suggest to replace $\hat{\sigma}^2$ with $V(k_{max}, \hat{F}^{k_{max}}) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2$, where $\hat{e}_{it} = X_{it} - \frac{1}{\sqrt{k}} \hat{\lambda}_i^{k_{max}} \hat{F}_t^{k_{max}}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$, the residuals for the linear regression of X on $\hat{F}^{k_{max}}$. Thus, the number of factors estimated using these three criteria may be sensitive to the selection of k_{max} . We will propose a method that avoids the sensitivity of selected \hat{k} depending on k_{max} .

Corollary 3.1 suggests the following three IC_p criteria can also be used to select the number of factors:

$$\begin{aligned} IC_{p1}(k) &= \ln(V(k, \hat{F}^k)) + k \cdot \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right), \\ IC_{p2}(k) &= \ln(V(k, \hat{F}^k)) + k \cdot \left(\frac{N+T}{NT}\right) \ln(\min\{N, T\}), \\ IC_{p3}(k) &= \ln(V(k, \hat{F}^k)) + k \cdot \left(\frac{\ln(\min\{N, T\})}{\min\{N, T\}}\right). \end{aligned} \tag{8}$$

The main advantage of the three criteria given in (8) is that the scaling factor $\hat{\sigma}^2$ is automatically removed by the logarithmic transformation. We do not need to estimate σ^2 before selecting the number of factors. Therefore, the number of factors estimated using IC_p criteria is insensitive to the selection of k_{max} .

As the estimated \hat{k} using PC_p criteria may be sensitive to k_{max} , the selection of k_{max} is an important issue in practice. Bai and Ng (2002) suggest to select k_{max} by setting $k_{max} = 8[(\min\{N, T\}/100)^{1/4}]$ where $[A]$ denotes the integer part of a real number A . But their theoretical result does not cover this case as this k_{max} increases (without bound) with N and T . Using some ad-hoc rules to select k_{max} may lead to $k_{max} < r$, which will lead to an underestimation

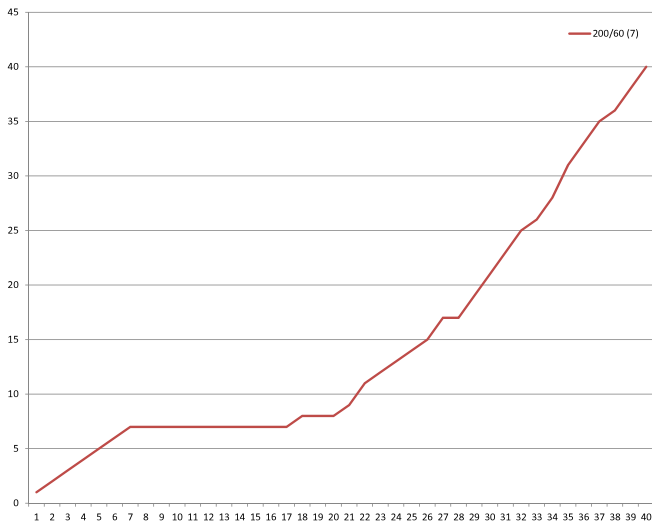


Fig. 1. Sensitivity of PC_{p1} Criterion to k_{max} : 200/60 case. Note: The values of \hat{k} estimated by PC_{p1} for $N = 200$, $T = 60$ and $r = 7$ with $k_{max} \in [1, 40]$.

of the number of factors because if $k_{max} < r$, then we will have $\hat{k} \leq k_{max} < r$. On the other hand, if k_{max} is too large ($k_{max} \gg r$), simulations show that the selected \hat{k} tends to overestimate r ($\hat{k} > r$). We propose a new procedure to resolve this problem. We propose to let k_{max} take a wide range of values. For each value of k_{max} , we select a $\hat{k}_{k_{max}}$ that minimizes the PC_p criteria. We then select the value of \hat{k} that appears most times among the different $\hat{k}_{k_{max}}$ values, i.e., we select the mode of $\hat{k}_{k_{max}}$ (over a wide range of k_{max}). We use a specific example to illustrate this selection procedure. We generate a simulated data of $N = 200$, $T = 60$ with the true number of factors $r = 7$. We let k_{max} take values from $\{1, 2, \dots, 40\}$. For each different $1 \leq k_{max} \leq 40$, we select a $\hat{k}_{k_{max}}$ by minimizing PC_{p1} criterion. The result is presented in Fig. 1. From Fig. 1 we observe that when $k_{max} < r = 7$, we select $\hat{k} = k_{max} < 7$ as expected; when $7 \leq k_{max} \leq 16$, we select $\hat{k} = 7$; when $k_{max} > 16$, the selected $\hat{k} > 7$. Moreover, \hat{k} increases with k_{max} . We also notice that $\hat{k} = 7$ is selected ten times (when $k_{max} = 7, 8, \dots, 16$), while all the other values are chosen no more than three times. For example, when $17 \leq k_{max} \leq 19$, the selected $\hat{k}_{k_{max}} = 8$, i.e., $\hat{k}_{k_{max}} = 8$ is selected three times. According to our selection rule, $\hat{k} = 7$ is selected because $\hat{k} = 7$ appears most times (10 times).

Fig. 2 plots $\hat{k}-k_{max}$ curves for different N , T and r values. We see that although \hat{k} increases with k_{max} for most cases, our proposed procedure can select the correct number of factors because $\hat{k}_{k_{max}}$ takes value r more often than taking any other values for all cases reported in Fig. 2. Hence, our proposed procedure of selecting \hat{k} is not sensitive to k_{max} provided that one let k_{max} take a wide range of values. Therefore, we suggest letting k_{max} take values in $\{1, 2, \dots, [6 \log(\max(N, T))]\}$ where $[A]$ denotes the integer part of a real number A . $[6 \log(\max(N, T))]$ is around 41, 45 and 55 when $\max(N, T) = 1000, 2000$ and 10000 . This setting is also consistent with our simulation since we let $r = [1.5 \log(\max(N, T))]$ in our simulations in Section 4.

4. Simulations

In this section we conduct Monte Carlo simulations to investigate how our modified criteria of Bai and Ng (2002) perform when the number of factors is allowed to increase with N or T . For simplicity of the comparison with the simulation results in Bai

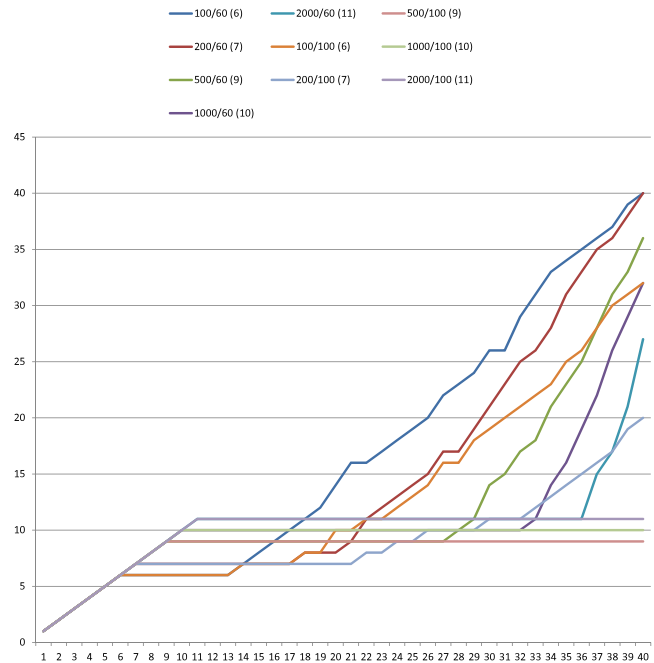


Fig. 2. Sensitivity of PC_{p1} Criterion to k_{max} . Note: Each line represents \hat{k} estimated by PC_{p1} for each case of different sample size. The notation in the graph shows the sample size and the true number of factors for each case. For example, 100/60(6) means that $N = 100$, $T = 60$ and $r = 6$.

and Ng (2002), we first fix T and allow N and r to increase. When T is fixed as 60, we let $N = 100, 200, 500, 1000, 2000$ and $r = [1.5 \log(N)]$, where $[A]$ denotes the integer part of a real number A ; for $T = 100$, we let $N = 40, 60, 100, 200, 500, 1000, 2000$ and $r = [1.5 \log(N)]$. The simulation results for this case are reported in the upper part of each table for each data generating process (DGP). Next, we check the performance of the criteria when N is fixed and T keeps increasing. When $N = 100$, we let $T = 40, 60, 100, 200, 500, 1000, 2000$ and $r = [1.5 \log(T)]$; when $N = 60$, we let $T = 100, 200, 500, 1000, 2000$ and $r = [1.5 \log(T)]$. The simulation results for this case are reported in the lower part of each table for each DGP. We replicate the suggested estimating procedure 1000 times and the reported results are the averages of \hat{k} over 1000 replications.

The data generating processes (DGP) have the following form:

$$X_{it} = \frac{1}{\sqrt{r}} \sum_{j=1}^r \lambda_{ij} F_{tj} + e_{it},$$

where $\lambda_{ij} \sim i.i.d. N(0, 1)$, $F_{tj} \sim i.i.d. N(0, 2)$.

We consider three DGPs here. In the base case, we set the DGP as $e_{it} \sim i.i.d.N(0, 1)$. This base DGP is denoted as DGP1. The simulation results for this case are reported in Table 1. We see that all information criteria give precise estimates of the number of factors.

For the heterogeneity case of DGP2, we set the idiosyncratic shocks to be heterogeneous. We let $e_{it} = u_{it} + \delta_t \epsilon_{it}$ where $u_{it} \sim i.i.d.N(0, 1)$, $\epsilon_{it} \sim i.i.d.N(0, 1)$, and $\delta_t = 0$ for even t , $\delta_t = 1$ for odd t . Thus the variance of the idiosyncratic shocks is 1 when t is odd and 2 when t is even. We denote this DGP as DGP2. The estimated values of \hat{k} are reported in Table 2 where the boldfaced numbers indicate incorrect selection of the number of factors. Similar to the homogeneous cases, PC_{p1} , PC_{p2} , and PC_{p3} perform well under all kinds of combinations of N and T . The other three criteria IC_{p1} and IC_{p2} , and IC_{p3} also perform well in general, although occasionally they may select \hat{k} that is slightly smaller than the true number of factors r when sample size is small.

Table 1
Estimated number of factors: DGP1.

<i>N</i>	<i>T</i>	<i>r</i>	<i>PC_{p1}</i>	<i>PC_{p2}</i>	<i>PC_{p3}</i>	<i>IC_{p1}</i>	<i>IC_{p2}</i>	<i>IC_{p3}</i>
100	60	6	6	6	6	6	6	6
200	60	7	7	7	7	7	7	7
500	60	9	9	9	9	9	9	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	5	5	5	5	5	5
60	100	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	5	5	5	5	5	5
100	60	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
100	200	7	7	7	7	7	7	7
100	500	9	9	9	9	9	9	9
100	1000	10	10	10	10	10	10	10
100	2000	11	11	11	11	11	11	11
60	100	6	6	6	6	6	6	6
60	200	7	7	7	7	7	7	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP1: $X_{it} = \frac{1}{\sqrt{T}} \sum_{j=1}^r \lambda_{ij} F_{jt} + e_{it}$; $r = [c * \ln(N)]$ for the upper part of the table, and $r = [c * \ln(T)]$ for the lower part, where $c = 1.5$, and $[A]$ denotes the integer part of a real number A .

Table 2
Estimated number of factors: Heterogeneity.

<i>N</i>	<i>T</i>	<i>r</i>	<i>PC_{p1}</i>	<i>PC_{p2}</i>	<i>PC_{p3}</i>	<i>IC_{p1}</i>	<i>IC_{p2}</i>	<i>IC_{p3}</i>
100	60	6	6	6	6	5	5	6
200	60	7	7	7	7	7	7	7
500	60	9	9	9	9	8	8	8
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	5	5	5	5	5	5
60	100	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	5	5	5	5	5	5
100	60	6	6	6	6	5	5	6
100	100	6	6	6	6	6	6	6
100	200	7	7	7	7	7	7	7
100	500	9	9	9	9	9	9	9
100	1000	10	10	10	10	10	10	10
100	2000	11	11	11	11	11	11	11
60	100	6	6	6	6	6	6	6
60	200	7	7	7	7	7	7	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP2: $X_{it} = \frac{1}{\sqrt{T}} \sum_{j=1}^r \lambda_{ij} F_{jt} + e_{it}$; $e_{it} = u_{it} + \delta_t \epsilon_{it}$, where $\delta_t = 0$ for t even, and $\delta_t = 1$ for t odd; $r = [c \ln(N)]$ for the upper part of the table, and $r = [c \ln(T)]$ for the lower part, where $[A]$ denotes taking the integer part of a real number.

For the last case, denoted as DGP3, we allow the idiosyncratic to be autocorrelated. We set $e_{it} = \rho e_{it-1} + v_{it}$, where $\rho = 0.5$ and $v_{it} \sim i.i.d.N(0, 1)$. The estimation results are reported in Table 3. The results for this case are almost the same as those of the base case except for $(N, T) = (60, 200)$ with $r = 7$, IC_{p1} and IC_{p2} select $r = 6$. All other four information criteria perform quite well in accurately estimating the number of factors for all (N, T) combinations for DGP3.

Summarizing the results for all the DGPs we observe that PC_{p1} , PC_{p2} , and PC_{p3} have the best overall performance. IC_{p1} , IC_{p2} , and IC_{p3} perform well when the sample size is large ($\min\{N, T\} > 100$).

Table 3
Estimated number of factors: Autocorrelation.

<i>N</i>	<i>T</i>	<i>r</i>	<i>PC_{p1}</i>	<i>PC_{p2}</i>	<i>PC_{p3}</i>	<i>IC_{p1}</i>	<i>IC_{p2}</i>	<i>IC_{p3}</i>
100	60	6	6	6	6	6	6	6
200	60	7	7	7	7	7	7	7
500	60	9	9	9	9	9	9	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	5	5	5	5	5	5
60	100	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	5	5	5	5	5	5
100	60	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
100	200	7	7	7	7	7	7	7
100	500	9	9	9	9	9	9	9
100	1000	10	10	10	10	10	10	10
100	2000	11	11	11	11	11	11	11
60	100	6	6	6	6	6	6	6
60	200	7	7	7	7	6	6	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP3: $X_{it} = \frac{1}{\sqrt{T}} \sum_{j=1}^r \lambda_{ij} F_{jt} + e_{it}$; $e_{it} = \rho e_{it-1} + v_{it}$; $\rho = 0.5$; $r = [c \ln(N)]$ for the upper part of the table, and $r = [c \ln(T)]$ for the lower part, where $[A]$ denotes taking the integer part of a real number.

5. Concluding remarks

In this paper, we consider the problem of determining the number of factors in large factor models where the number of factors is allowed to increase, but with a slower rate, as N or T increases. We extend the analysis of Bai and Ng (2002) to the case that number of factors can increase with the sample size and prove the consistency of a modified Bai and Ng's (2002) procedure in determining the number of factors. We also propose a ('mode' based) new procedure so that our selected number of factors is not sensitive to the choice of k_{max} . Monte Carlo simulation results suggest that the criteria PC_{p1} , PC_{p2} and PC_{p3} all have the overall best performance. Other criteria such as IC_{p1} , IC_{p2} and PC_{p3} can also be used to accurately estimate the number of factors when the data dimensions are relatively large, say $\min\{N, T\} \geq 100$. One possible future research topic is to find alternative criteria that can improve the finite-sample performance of Bai and Ng's (2002) procedure and our modified procedure such that the new criteria can accurately determine the number of factors even in small or medium size samples.

Acknowledgments

We would like to thank two anonymous referees, an associate editor and seminar participants at Princeton University, University of Southern California, University of Kansas, Xiamen University and Texas A&M University for helpful comments that greatly improved our paper. Hongjun Li's research is partially supported by China National Science Foundation, project # 71601130.

Appendix A. Proofs

Proof of Proposition 1

We will first prove a lemma (Lemma 1) below which will be used in proving Proposition 1.

Lemma 1. Under Assumptions A–C, we have for some positive constant $0 < M_2 < \infty$, and for all N and T ,

- (i) $\frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T \gamma_N(s, t)^2 \leq M_2$;
(ii) $E \left(\frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^T \left(\frac{1}{N} \sum_{i=1}^N X_{it} X_{is} \right)^2 \right) \leq M_2$.

Proof. (i) See the proof of Lemma 1(i) in Bai and Ng (2002).

- (ii) It suffices to prove that for all (i, t) that $E(X_{it}^4) \leq M$. Now $E(X_{it}^4) \leq 8r^{-2}E[(\lambda_i^0 F_t^0)^4] + 8E(e_{it}^4) \leq 16M_1$ by Assumptions A5 and B1. \square

Proof of Proposition 3.1. Recall that $\hat{F}^k = \frac{\sqrt{k}}{N} X \tilde{\Lambda}^k$ and $\tilde{\Lambda}^k = \frac{\sqrt{k}}{T} X' \tilde{F}^k$. From the normalization $\tilde{F}^k \tilde{F}^k / T = I_k$, we also have $(Tk)^{-1} \sum_{t=1}^T \|\tilde{F}_t^k\|^2 = 1$. Following Bai and Ng (2002), using $H^k = (\tilde{F}^k F^0 / T)(\Lambda^0 / \Lambda^0 / N)$, we have

$$\begin{aligned} \hat{F}_t^k - H^k F_t^0 &= \frac{k}{T} \sum_{s=1}^T \tilde{F}_s^k \gamma_N(s, t) + \frac{k}{T} \sum_{s=1}^T \tilde{F}_s^k \zeta_{st} \\ &+ \frac{k}{T} \sum_{s=1}^T \tilde{F}_s^k \eta_{st} + \frac{k}{T} \sum_{s=1}^T \tilde{F}_s^k \xi_{st}, \end{aligned}$$

where $\zeta_{st} = e'_s e_t / N - \gamma_N(s, t)$, $\eta_{st} = F_s^0 \Lambda^0 e_t / (N\sqrt{r})$, and $\xi_{st} = F_t^0 \Lambda^0 e_s / (N\sqrt{r}) = \eta_{ts}$.

Because $(x + y + z + u)^2 \leq 4(x^2 + y^2 + z^2 + u^2)$, $\|\hat{F}_t^k - H^k F_t^0\|^2 \leq 4(a_t + b_t + c_t + d_t)$, where $a_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \gamma_N(s, t) \right\|^2$, $b_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \zeta_{st} \right\|^2$, $c_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \eta_{st} \right\|^2$ and $d_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \xi_{st} \right\|^2$. It follows that $(1/T) \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \leq (4/T) \sum_{t=1}^T (a_t + b_t + c_t + d_t)$.

By Cauchy's inequality, we have $\left\| \sum_{s=1}^T \tilde{F}_s^k \gamma_N(s, t) \right\|^2 \leq \left(\sum_{s=1}^T \|\tilde{F}_s^k\|^2 \right) \cdot \left(\sum_{s=1}^T \gamma_N(s, t)^2 \right)$. Thus,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T a_t &\leq \frac{kk^2}{T} \left(\frac{1}{Tk} \sum_{s=1}^T \|\tilde{F}_s^k\|^2 \right) \cdot \frac{1}{T} \left(\sum_{t=1}^T \sum_{s=1}^T \gamma_N(s, t)^2 \right) \\ &= O_p \left(\frac{k^3}{T} \right) \end{aligned}$$

by Lemma 1(i) and the fact that $(Tk)^{-1} \sum_{t=1}^T \|\tilde{F}_t^k\|^2 = 1$ (this follows from $\tilde{F}^k \tilde{F}^k / T = I_k$).

For b_t , we have that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T b_t &= \frac{k^2}{T^3} \sum_{t=1}^T \left\| \sum_{s=1}^T \tilde{F}_s^k \zeta_{st} \right\|^2 \\ &= \frac{k^2}{T^3} \sum_{t=1}^T \sum_{s=1}^T \sum_{u=1}^T \tilde{F}_s^k \tilde{F}_u^k \zeta_{st} \zeta_{ut} \\ &\leq \frac{k^2}{T} \left(\frac{1}{T^2} \sum_{s=1}^T \sum_{u=1}^T (\tilde{F}_s^k \tilde{F}_u^k)^2 \right)^{1/2} \left[\frac{1}{T^2} \sum_{s=1}^T \sum_{u=1}^T \left(\sum_{t=1}^T \zeta_{st} \zeta_{ut} \right)^2 \right]^{1/2} \\ &\leq \frac{k^3}{T} \left(\frac{1}{Tk} \sum_{s=1}^T \|\tilde{F}_s^k\|^2 \right) \left[\frac{1}{T^2} \sum_{s=1}^T \sum_{u=1}^T \left(\sum_{t=1}^T \zeta_{st} \zeta_{ut} \right)^2 \right]^{1/2} \\ &= k^3 \left[\frac{1}{T^4} \sum_{s=1}^T \sum_{u=1}^T \left(\sum_{t=1}^T \zeta_{st} \zeta_{ut} \right)^2 \right]^{1/2} \\ &= O_p \left(\frac{k^3}{N} \right), \end{aligned}$$

where the last equality follows from $\left[\frac{1}{T^4} \sum_{s=1}^T \sum_{u=1}^T \left(\sum_{t=1}^T \zeta_{st} \zeta_{ut} \right)^2 \right]^{1/2} = O_p(N^{-1})$ as shown in the proof of Theorem 1 of Bai and Ng (2002).

From $E(T^{-1} \sum_{t=1}^T \zeta_{st} \zeta_{ut})^2 = E(T^{-2} \sum_{t=1}^T \sum_{v=1}^T \zeta_{st} \zeta_{ut} \zeta_{sv} \zeta_{uv}) \leq \max_{s,t} E|\zeta_{st}|^4$ and

$$E|\zeta_{st}|^4 = \frac{1}{N^2} E \left| \frac{1}{\sqrt{N}} \sum_{i=1}^N (e_{it} e_{is} - E(e_{it} e_{is})) \right|^4 \leq \frac{1}{N^2} M_1$$

by Assumption B5, we have

$$\frac{1}{T} \sum_{t=1}^T b_t \leq O_p(k^3) \frac{1}{T} \sqrt{\frac{T^2}{N^2}} = O_p \left(\frac{k^3}{N} \right).$$

For c_t , we have

$$\begin{aligned} c_t &= \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \eta_{st} \right\|^2 \\ &= \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k F_s^0 \Lambda^0 e_t / N\sqrt{r} \right\|^2 \\ &\leq \frac{k^2}{N^2} \|e'_t \Lambda^0 / \sqrt{r}\|^2 \left(\frac{k}{Tk} \sum_{s=1}^T \|\tilde{F}_s^k\|^2 \right) \left(\frac{r}{Tr} \sum_{s=1}^T \|F_s^0\|^2 \right) \\ &= \frac{k^2}{N^2} \|e'_t \Lambda^0 / \sqrt{r}\|^2 O_p(kr) \end{aligned}$$

because $\frac{1}{Tk} \sum_{s=1}^T \|\tilde{F}_s^k\|^2 = 1$ and $\frac{1}{Tr} \sum_{s=1}^T \|F_s^0\|^2 = O_p(1)$. It follows that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T c_t &= O_p(kr) \frac{k^2}{N} \frac{1}{T} \sum_{t=1}^T \left\| \frac{e'_t \Lambda^0}{\sqrt{Nr}} \right\|^2 \\ &= O_p \left(\frac{k^3 r}{N} \right) \end{aligned}$$

because $\frac{1}{T} \sum_{t=1}^T \left\| \frac{e'_t \Lambda^0}{\sqrt{Nr}} \right\|^2 = O_p(1)$ by Assumption C2.

The term $(1/T) \sum_{t=1}^T d_t = O_p \left(\frac{k^3 r}{N} \right)$ can be proved similarly. Combining the above results, we have shown that

$$\begin{aligned} (1/T) \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 &\leq (4/T) \sum_{t=1}^T (a_t + b_t + c_t + d_t) \\ &= O_p \left(\frac{k^3 r}{N} \right) + O_p \left(\frac{k^3}{T} \right). \end{aligned}$$

Alternatively, Proposition 3.1 can be proved by concentrating over F_t . Following the similar steps, we can show that

$$\begin{aligned} (1/N) \sum_{i=1}^N \|\hat{\lambda}_i^k - H^k \lambda_i^0\|^2 &\leq (4/N) \sum_{i=1}^N (a_i + b_i + c_i + d_i) \\ &= O_p \left(\frac{k^3}{N} \right) + O_p \left(\frac{k^3 r}{T} \right). \quad \square \end{aligned}$$

Proof of Proposition 2

Proof. From the proof of Proposition 1 we know that

$$\frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \leq (4/T) \sum_{t=1}^T (a_t + b_t + c_t + d_t)$$

and that $T^{-1} \sum_{t=1}^T a_t = O_p(k^3/T)$ and $T^{-1} \sum_{t=1}^T d_t = O_p(k^3/N)$. Therefore, we only need to show that $T^{-1} \sum_{t=1}^T c_t = O_p(k^3/N)$ and $T^{-1} \sum_{t=1}^T d_t = O_p(k^3/N)$. Since the proofs are similar, we will only prove for the term related to c_t .

For c_t , we have

$$\begin{aligned} c_t &= \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k F_s^{0'} \Lambda^{0'} e_t / N \sqrt{r} \right\|^2 \\ &\leq \frac{k^2}{N} \frac{1}{r} \left(\frac{1}{T} \sum_{s=1}^T \|\tilde{F}_s^k\|^2 \right) \left(\frac{1}{TN} \sum_{s=1}^T \|F_s^{0'} \Lambda^{0'} e_t\|^2 \right) \\ &= O_p(k^3/N) \frac{1}{r} \left(\frac{1}{TN} \sum_{s=1}^T \|F_s^{0'} \Lambda^{0'} e_t\|^2 \right) \end{aligned}$$

because $T^{-1} \sum_{s=1}^T \|\tilde{F}_s^k\|^2 = O_p(k)$.

Next, we show that $A \stackrel{\text{def}}{=} (TN)^{-1} \sum_{s=1}^T \|F_s^{0'} \Lambda^{0'} e_t\|^2 = O_p(r)$.

$$\begin{aligned} E(|A|) &= \frac{1}{NT} \sum_{s=1}^T E \|F_s^{0'} \Lambda^{0'} e_t\|^2 \\ &= \frac{1}{NT} \sum_{s=1}^T \sum_{l=1}^r \sum_{m=1}^r \sum_{i=1}^N \sum_{j=1}^N E(e_{it} e_{jt}) E(F_{sl}^0 F_{sm}^0) \Lambda_{il}^0 \lambda_{jm}^0 \\ &= \frac{1}{NT} \sum_{s=1}^T \sum_{l=1}^r \sum_{i=1}^N E(e_{it}^2) E((F_{sl}^0)^2) (\Lambda_{il}^0)^2 \\ &= O(r), \end{aligned}$$

because of the zero correlation assumptions that $E(e_{it} e_{jt}) = 0$ for $j \neq i$ and $E(F_{sl}^0 F_{sm}^0) = 0$ for $m \neq l$. This implies that $A = O_p(r)$. Hence, $c_t = O_p(k^3/N)$. This completes the proof of Proposition 2. \square

From the above proof we can see that the conclusion of Proposition 2 still holds true if the zero correlation assumptions are replaced by some weakly dependent assumptions such as $N^{-1} \sum_{i=1}^N \sum_{j \neq i}^N E(e_{it} e_{jt} c_{1,ijlm}) = O(1)$ and $r^{-1} \sum_{l=1}^r \sum_{m \neq l}^r E(F_{sl}^0 F_{sm}^0 c_{2,lmim}) = O(1)$, where $c_{1,ijlm}$ and $c_{2,ijlm}$ are some bounded sequences of non-random numbers depending on i, j, l, m .

Proof of Theorem 3.1

Lemma 2. Let $D_k = \hat{F}^k \hat{F}^k / T$ and $D_0 = H^k F^{0'} F^0 H^k / T$. When $k \leq r$, we have (i) $\|D_k^{-1}\| = O_p(k)$; (ii) $\|D_k^{-1} - D_0^{-1}\| = O_p\left(\max\left\{\frac{k_{\max}^4 r^{1.5}}{\sqrt{N}}, \frac{k_{\max}^4 r}{\sqrt{T}}\right\}\right)$.

Proof. Following Bai and Ng (2002), we have

$$\begin{aligned} D_k - D_0 &= \frac{\hat{F}^k \hat{F}^k}{T} - \frac{H^k F^{0'} F^0 H^k}{T} \\ &= \frac{1}{T} \sum_{t=1}^T [\hat{F}_t^k \hat{F}_t^k - H^k F_t^0 F_t^0 H^k] \\ &= \frac{1}{T} \sum_{t=1}^T (\hat{F}_t^k - H^k F_t^0) (\hat{F}_t^k - H^k F_t^0)' \\ &\quad + \frac{1}{T} \sum_{t=1}^T (\hat{F}_t^k - H^k F_t^0) F_t^0 H^k \\ &\quad + \frac{1}{T} \sum_{t=1}^T H^k F_t^0 (\hat{F}_t^k - H^k F_t^0)'. \end{aligned}$$

Hence, we have

$$\begin{aligned} \|D_k - D_0\| &\leq \frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \\ &\quad + 2 \left(\frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \|H^k F_t^0\|^2 \right)^{1/2} \\ &= O_p\left(\max\left\{\frac{k^3 r}{N}, \frac{k^3}{T}\right\}\right) \\ &\quad + O_p\left(\max\left\{\frac{\sqrt{k^3 r}}{\sqrt{N}}, \frac{\sqrt{k^3}}{\sqrt{T}}\right\}\right) \cdot O_p(\sqrt{kr^2}) \\ &= O_p\left(\max\left\{\frac{k^2 r^{1.5}}{\sqrt{N}}, \frac{k^2 r}{\sqrt{T}}\right\}\right) \end{aligned}$$

by Proposition 3.1 and the fact that $\frac{1}{T} \sum_{t=1}^T \|H^k F_t^0\|^2 = O_p(kr^2)$, which is shown below.

From weakly dependent process of F_t^0 , it is easy to show that

$$\frac{1}{T} \sum_{t=1}^T \|H^k F_t^0\|^2 - E \left[\frac{1}{T} \sum_{t=1}^T \|H^k F_t^0\|^2 \right] = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Also, one can easily show that $\|D_k^{-1}\| = O_p(k)$. Then from $D_k^{-1} - D_0^{-1} = D_k^{-1} (D_0 - D_k) D_0^{-1}$, we have

$$\begin{aligned} \|D_k^{-1} - D_0^{-1}\| &= \|D_k^{-1} (D_0 - D_k) D_0^{-1}\| \\ &\leq \|D_k^{-1}\| \cdot \|D_0 - D_k\| \cdot \|D_0^{-1}\| \\ &= k^2 \frac{\|D_k^{-1}\|}{k} \cdot \|D_0 - D_k\| \cdot \frac{\|D_0^{-1}\|}{k} \\ &= k^2 \cdot O_p(1) \cdot O_p\left(\max\left\{\frac{k^2 r^{1.5}}{\sqrt{N}}, \frac{k^2 r}{\sqrt{T}}\right\}\right) \\ &= O_p\left(\max\left\{\frac{k_{\max}^4 r^{1.5}}{\sqrt{N}}, \frac{k_{\max}^4 r}{\sqrt{T}}\right\}\right). \quad \square \end{aligned}$$

Lemma 3. For $1 \leq k \leq r$, and the H^k defined in Proposition 3.1, we have

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = O_p\left(\max\left\{\frac{k_{\max}^5 r^{3.5}}{\sqrt{N}}, \frac{k_{\max}^5 r^3}{\sqrt{T}}\right\}\right).$$

Proof. For the true factor matrix with r factors and H^k defined in Proposition 3.1, let $M_{FH}^0 = I - P_{FH}^0$ denote the idempotent matrix spanned by null space of $F^0 H^k$, with $P_{FH}^0 = F^0 H^k (H^k F^{0'} F^0 H^k)^{-1} H^k F^{0'}$. Correspondingly, let $M_{\hat{F}}^k = I_T - \hat{F}^k (\hat{F}^k \hat{F}^k)^{-1} \hat{F}^k = I_T - P_{\hat{F}}^k$. Then

$$\begin{aligned} V(k, \hat{F}^k) &= \frac{1}{NT} \sum_{i=1}^N X_i' M_{\hat{F}}^k X_i, \\ V(k, F^0 H^k) &= \frac{1}{NT} \sum_{i=1}^N X_i' M_{FH}^0 X_i, \\ V(k, \hat{F}^k) - V(k, F^0 H^k) &= \frac{1}{NT} \sum_{i=1}^N X_i' (P_{FH}^0 - P_{\hat{F}}^k) X_i. \end{aligned}$$

Following Bai and Ng (2002), let $D_k = \hat{F}^k \hat{F}^k / T$ and $D_0 = H^k F^{0'} F^0 H^k / T$. Then

$$P_{\hat{F}}^k - P_{FH}^0 = \frac{1}{T} \hat{F}^k \left(\frac{\hat{F}^k \hat{F}^k}{T} \right)^{-1} \hat{F}^k - \frac{1}{T} F^0 H^k \left(\frac{H^k F^{0'} F^0 H^k}{T} \right)^{-1} H^k F^{0'}$$

$$\begin{aligned} &= \frac{1}{T} [\hat{F}^k D_k^{-1} \hat{F}^k - F^0 H^k D_0^{-1} H^k F^0] \\ &= \frac{1}{T} [(\hat{F}^k - F^0 H^k + F^0 H^k) D_k^{-1} (\hat{F}^k - F^0 H^k + F^0 H^k)' \\ &\quad - F^0 H^k D_0^{-1} H^k F^0] \\ &= \frac{1}{T} [(\hat{F}^k - F^0 H^k) D_k^{-1} (\hat{F}^k - F^0 H^k)' + (\hat{F}^k - F^0 H^k) D_k^{-1} H^k F^0 \\ &\quad + F^0 H^k D_k^{-1} (\hat{F}^k - F^0 H^k)' - F^0 H^k D_0^{-1} H^k F^0]. \end{aligned}$$

Thus, $N^{-1} T^{-1} \sum_{i=1}^N \underline{X}_i' (P_{\hat{F}}^k - P_{FH}^0) \underline{X}_i = I + II + III + IV$. We consider each term in turn.

$$\begin{aligned} I &= \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T (\hat{F}_t^k - H^k F_t^0)' D_k^{-1} (\hat{F}_s^k - H^k F_s^0) X_{it} X_{is} \\ &\leq \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T (\hat{F}_t^k - H^k F_t^0)' D_k^{-1} (\hat{F}_s^k - H^k F_s^0) \right)^{1/2} \\ &\quad \cdot \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left(\frac{1}{N} \sum_{i=1}^N X_{it} X_{is} \right)^2 \right)^{1/2} \\ &\leq \left(\frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \right) \cdot \|D_k^{-1}\| \cdot O_p(1) \\ &= O_p \left(\max \left\{ \frac{k^3 r}{N}, \frac{k^3}{T} \right\} \right) \cdot k \cdot O_p(1) \\ &= O_p \left(\max \left\{ \frac{k^4 r}{N}, \frac{k^4}{T} \right\} \right) \end{aligned}$$

by Proposition 3.1, Lemma 1(iii) and Lemma 2(i).

$$\begin{aligned} II &= \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T (\hat{F}_t^k - H^k F_t^0)' D_k^{-1} H^k F_s^0 X_{it} X_{is} \\ &\leq \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \cdot \|H^k F_s^0\|^2 \cdot \|D_k^{-1}\|^2 \right)^{1/2} \\ &\quad \cdot \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left(\frac{1}{N} \sum_{i=1}^N X_{it} X_{is} \right)^2 \right)^{1/2} \\ &\leq \left(\frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \right)^{1/2} \cdot \|D_k^{-1}\| \\ &\quad \cdot \left(\frac{kr^2}{Tkr^2} \sum_{s=1}^T \|H^k F_s^0\|^2 \right)^{1/2} \cdot O_p(1) \\ &= O_p \left(\max \left\{ \left(\frac{k^3 r}{N} \right)^{1/2}, \left(\frac{k^3}{T} \right)^{1/2} \right\} \right) \cdot k \cdot k^{1/2} r \cdot O_p(1) \\ &= O_p \left(\max \left\{ \frac{k^3 r^{1.5}}{\sqrt{N}}, \frac{k^3 r}{\sqrt{T}} \right\} \right). \end{aligned}$$

Similarly, one can verify that III is also $O_p \left(\max \left\{ \frac{k^3 r^{1.5}}{\sqrt{N}}, \frac{k^3 r}{\sqrt{T}} \right\} \right)$.

$$\begin{aligned} IV &= \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T F_t^0 H^k (D_k^{-1} - D_0^{-1}) H^k F_s^0 X_{it} X_{is} \\ &\leq \|D_k^{-1} - D_0^{-1}\| \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T \|H^k F_t^0\| \cdot |X_{it}| \right)^2 \end{aligned}$$

$$\begin{aligned} &\leq \|D_k^{-1} - D_0^{-1}\| \frac{kr^2}{N} \sum_{i=1}^N \left(\frac{1}{T\sqrt{kr}} \sum_{t=1}^T \|H^k F_t^0\| \right)^2 \\ &= \|D_k^{-1} - D_0^{-1}\| \cdot kr^2 \cdot O_p(1) \\ &= O_p \left(\max \left\{ \frac{k^4 r^{4.5}}{\sqrt{N}}, \frac{k^4 r^4}{\sqrt{T}} \right\} \right), \end{aligned}$$

where we used $\|D_k^{-1} - D_0^{-1}\| = O_p \left(\max \left\{ \frac{k^3 r^{2.5}}{\sqrt{N}}, \frac{k^3 r^2}{\sqrt{T}} \right\} \right)$ by Lemma 2(ii).

Thus, we have

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = O_p \left(\max \left\{ \frac{k_{\max}^5 r^{3.5}}{\sqrt{N}}, \frac{k_{\max}^5 r^3}{\sqrt{T}} \right\} \right). \quad \square$$

Lemma 4. For the matrix H^k defined in Proposition 3.1, and for each k with $k < r = r_{N,T} \rightarrow \infty$, there exists a positive constant C such that

$$\text{plim}_{N,T \rightarrow \infty} \inf_k [V(k, F^0 H^k) - V(r, F^0)] \geq C > 0.$$

Proof.

$$\begin{aligned} V(k, F^0 H^k) - V(r, F^0) &= \frac{1}{NT} \sum_{i=1}^N \underline{X}_i' M_{FH}^0 \underline{X}_i - \frac{1}{NT} \sum_{i=1}^N \underline{X}_i' M_F^0 \underline{X}_i \\ &= \frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{\sqrt{r}} F^0 \lambda_i^0 + \underline{e}_i \right)' M_{FH}^0 \left(\frac{1}{\sqrt{r}} F^0 \lambda_i^0 + \underline{e}_i \right) \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' M_F^0 \underline{e}_i \\ &= \frac{1}{NT r} \sum_{i=1}^N \lambda_i^0 F^0 M_{FH}^0 F^0 \lambda_i^0 + \frac{2}{NT \sqrt{r}} \sum_{i=1}^N \underline{e}_i' M_{FH}^0 F^0 \lambda_i^0 \\ &\quad + \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' (P_F^0 - P_{FH}^0) \underline{e}_i \\ &= A + B + D. \end{aligned}$$

Notice that $P_F^0 - P_{FH}^0 \geq 0$, thus $III \geq 0$. For the first term,

$$\begin{aligned} A &= \frac{1}{NTr} \sum_{i=1}^N \lambda_i^0 F^0 M_{FH}^0 F^0 \lambda_i^0 \\ &= \frac{1}{NTr} \sum_{i=1}^N (M_{FH}^0 F^0 \lambda_i^0)' M_{FH}^0 F^0 \lambda_i^0 \\ &\geq C > 0 \end{aligned}$$

because $k < r$ and $M_{FH}^0 F^0 \lambda_i^0 \neq 0$.

Next,

$$B = \frac{2}{NT \sqrt{r}} \sum_{i=1}^N \underline{e}_i' F^0 \lambda_i^0 - \frac{2}{NT \sqrt{r}} \sum_{i=1}^N \underline{e}_i' P_{FH}^0 F^0 \lambda_i^0.$$

Consider the first term

$$\begin{aligned} \left| \frac{1}{NT \sqrt{r}} \sum_{i=1}^N \underline{e}_i' F^0 \lambda_i^0 \right| &= \left| \frac{1}{NT \sqrt{r}} \sum_{i=1}^N \sum_{t=1}^T \underline{e}_{it} F_t^0 \lambda_i^0 \right| \\ &\leq \left(\frac{1}{Tr} \sum_{t=1}^T \|F_t^0\|^2 \right)^{1/2} \cdot \sqrt{r} \cdot \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \underline{e}_{it} \lambda_i^0 \right\|^2 \right)^{1/2} \\ &= O_p \left(\frac{\sqrt{r}}{\sqrt{N}} \right), \end{aligned}$$

where the last equality follows from [Assumption C2](#). The second term is also $o_p(1)$, and hence $B = o_p(1)$. \square

Lemma 5. For any k with $r \leq k \leq k_{max}$, $V(k, \hat{F}^k) - V(r, \hat{F}^r) = O_p\left(\max\left\{\frac{k_{max}^2 r}{N}, \frac{k_{max}^2 r^{0.5}}{T}\right\}\right)$.

Proof.

$$\begin{aligned} |V(k, \hat{F}^k) - V(r, \hat{F}^r)| &\leq |V(k, \hat{F}^k) - V(r, F^0)| \\ &\quad + |V(r, F^0) - V(r, \hat{F}^r)| \\ &\leq 2 \max_{r \leq k} |V(k, \hat{F}^k) - V(r, F^0)|. \end{aligned}$$

Thus, it is sufficient to prove for each k with $r \leq k \leq k_{max}$,

$$V(k, \hat{F}^k) - V(r, F^0) = O_p\left(\max\left\{\frac{k_{max} r^2}{\sqrt{N}}, \frac{k_{max} r^{1.5}}{\sqrt{T}}\right\}\right).$$

Let H^k be as defined in Proposition 3.1, with full row rank. Let the $k \times r$ matrix H^{k+} be the generalized inverse of H^k such that $H^k H^{k+} = I_r$. From $\underline{X}_i = \frac{1}{\sqrt{r}} F^0 \lambda_i^0 + \underline{e}_i$, we have $\underline{X}_i = \frac{1}{\sqrt{r}} F^0 H^k H^{k+} \lambda_i^0 + \underline{e}_i$. This implies that

$$\begin{aligned} \underline{X}_i &= \frac{1}{\sqrt{r}} \hat{F}^k H^{k+} \lambda_i^0 + \underline{e}_i - \frac{1}{\sqrt{r}} (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \\ &= \frac{1}{\sqrt{r}} \hat{F}^k H^{k+} \lambda_i^0 + \underline{u}_i, \end{aligned}$$

where $\underline{u}_i = \underline{e}_i - \frac{1}{\sqrt{r}} (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0$.

Note that

$$\begin{aligned} V(k, \hat{F}^k) &= \frac{1}{NT} \sum_{i=1}^N \underline{u}_i' M_{\hat{F}^k}^k \underline{u}_i, \\ V(r, F^0) &= \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' M_{F^0}^0 \underline{e}_i, \\ V(k, \hat{F}^k) &= \frac{1}{NT} \sum_{i=1}^N \left(\underline{e}_i - \frac{1}{\sqrt{r}} (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \right)' \\ &\quad \times M_{\hat{F}^k}^k \left(\underline{e}_i - \frac{1}{\sqrt{r}} (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \right), \\ &= \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' M_{\hat{F}^k}^k \underline{e}_i - \frac{2}{NT\sqrt{r}} \sum_{i=1}^N \lambda_i^0' H^{k+} (\hat{F}^k - F^0 H^k)' M_{\hat{F}^k}^k \underline{e}_i \\ &\quad + \frac{1}{NT r} \sum_{i=1}^N \lambda_i^0' H^{k+} (\hat{F}^k - F^0 H^k)' M_{\hat{F}^k}^k (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \\ &= a + b + c. \end{aligned}$$

Because $I - M_{\hat{F}^k}^k$ is positive semi-definite, $\underline{x}' M_{\hat{F}^k}^k \underline{x} \leq \underline{x}' \underline{x}$. Thus

$$\begin{aligned} c &\leq \frac{1}{NT r} \sum_{i=1}^N \lambda_i^0' H^{k+} (\hat{F}^k - F^0 H^k)' (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \\ &\leq \frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - H^k F_t^0\|^2 \cdot \left(\frac{1}{Nr} \sum_{i=1}^N \|\lambda_i^0\|^2 \|H^{k+}\|^2 \right) \\ &= O_p\left(\max\left\{\frac{k^3 r}{N}, \frac{k^3}{T}\right\}\right) \cdot O_p(kr) \\ &= O_p\left(\max\left\{\frac{k^4 r^2}{N}, \frac{k^4 r}{T}\right\}\right) \end{aligned}$$

by Proposition 3.1.

For term b , we use the fact that $|\text{tr}(A)| \leq r \|A\|$ for any $r \times r$ matrix A . Thus

$$\begin{aligned} b &= \frac{2}{T\sqrt{r}} \text{tr} \left(H^{k+} (\hat{F}^k - F^0 H^k)' M_{\hat{F}^k}^k \left(\frac{1}{N} \sum_{i=1}^N \underline{e}_i \lambda_i^0 \right) \right) \\ &\leq 2 \cdot \|H^{k+}\| \cdot \left\| \frac{\hat{F}^k - F^0 H^k}{\sqrt{T}} \right\| \cdot \left\| \frac{1}{\sqrt{Tr} N} \sum_{i=1}^N \underline{e}_i \lambda_i^0 \right\| \\ &\leq 2 \cdot \|H^{k+}\| \cdot \left(\frac{1}{T} \sum_{t=1}^T \|\hat{F}_t^k - F^0 H^k\|^2 \right)^{1/2} \\ &\quad \cdot \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{Nr}} \sum_{i=1}^N \underline{e}_i \lambda_i^0 \right\|^2 \right)^{1/2} \\ &= 2 \cdot (kr)^{1/2} \cdot O_p\left(\max\left\{\frac{\sqrt{k^3 r}}{\sqrt{N}}, \frac{\sqrt{k^3}}{\sqrt{T}}\right\}\right) \cdot O_p(1) \\ &= O_p\left(\max\left\{\frac{k^2 r}{\sqrt{N}}, \frac{k^2 r^{0.5}}{\sqrt{T}}\right\}\right) \end{aligned}$$

by Proposition 3.1 and [Assumption C2](#). Therefore,

$$V(k, \hat{F}^k) = \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' M_{\hat{F}^k}^k \underline{e}_i + O_p\left(\max\left\{\frac{k^2 r}{\sqrt{N}}, \frac{k^2 r^{0.5}}{\sqrt{T}}\right\}\right).$$

Thus we have

$$\begin{aligned} V(k, \hat{F}^k) - V(r, F^0) &= \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' P_{\hat{F}^k}^0 \underline{e}_i - \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' P_{\hat{F}^k}^k \underline{e}_i \\ &\quad + O_p\left(\max\left\{\frac{k^2 r}{\sqrt{N}}, \frac{k^2 r^{0.5}}{\sqrt{T}}\right\}\right). \end{aligned}$$

Note that

$$\begin{aligned} \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' P_{\hat{F}^k}^0 \underline{e}_i &\leq \left\| \left(\frac{F^0 F^0}{T} \right)^{-1} \right\| \cdot \frac{1}{NT^2} \sum_{i=1}^N \underline{e}_i' F^0 F^0 \underline{e}_i \\ &= \left\| \left(\frac{F^0 F^0}{T} \right)^{-1} \right\| \cdot \frac{1}{NT} \sum_{i=1}^N \left\| \frac{1}{\sqrt{Tr}} \sum_{t=1}^T F_t^0 \underline{e}_{it} \right\|^2 \cdot r \\ &= r \cdot O_p(1) \cdot \frac{1}{T} \cdot r \cdot O_p(1) \\ &= O_p\left(\frac{r^2}{T}\right) \leq O_p\left(\max\left\{\frac{k^2 r}{\sqrt{N}}, \frac{k^2 r^{0.5}}{\sqrt{T}}\right\}\right). \end{aligned}$$

$\frac{1}{NT} \sum_{i=1}^N \underline{e}_i' P_{\hat{F}^k}^0 \underline{e}_i$ is bounded by the sum of the first k largest eigenvalues of the matrix $A_{NT} = \frac{1}{NT} e' e$, where $e = (e_{it})$, $T \times N$. Let $\rho(A)$ denote the largest eigenvalue of a matrix A . Under [Assumption B6](#), as [Bai and Ng \(2002\)](#) show, $\rho(A_{NT}) = O_p(C_{NT}^{-2})$, where $C_{NT}^2 = \min\{N, T\}$. Thus,

$$\begin{aligned} \frac{1}{NT} \sum_{i=1}^N \underline{e}_i' P_{\hat{F}^k}^0 \underline{e}_i &= O_p\left(\max\left\{\frac{k}{N}, \frac{k}{T}\right\}\right) \\ &\leq O_p\left(\max\left\{\frac{k^2 r}{\sqrt{N}}, \frac{k^2 r^{0.5}}{\sqrt{T}}\right\}\right). \end{aligned}$$

In summary, we have shown that

$$V(k, \hat{F}^k) - V(r, F^0) = O_p\left(\max\left\{\frac{k_{max}^2 r}{\sqrt{N}}, \frac{k_{max}^2 r^{0.5}}{\sqrt{T}}\right\}\right). \quad \square$$

Proof of Theorem 3.1

Proof. We shall prove that $\lim_{N,T \rightarrow \infty} P(PC(k) < PC(r)) = 0$ for all $k \neq r$. Since

$$PC(k) - PC(r) = V(k, \hat{F}^k) - V(r, \hat{F}^r) - (r - k)g(N, T),$$

it is sufficient to prove that $P[V(k, \hat{F}^k) - V(r, \hat{F}^r) < (r - k)g(N, T)] \rightarrow 0$ as $N, T, k, r \rightarrow \infty$.

Consider $k < r$. We have the identity:

$$\begin{aligned} V(k, \hat{F}^k) - V(r, \hat{F}^r) &= [V(k, \hat{F}^k) - V(k, F^0 H^k)] \\ &\quad + [V(k, F^0 H^k) - V(r, F^0 H^r)] \\ &\quad + [V(r, F^0 H^r) - V(r, \hat{F}^r)]. \end{aligned}$$

Lemma 3 implies that the first and third terms are both $O_p\left(\max\left\{\frac{k_{\max}^{3.5}}{\sqrt{N}}, \frac{k_{\max}^8}{\sqrt{T}}\right\}\right)$. Next, we consider the second item. Because $F^0 H^r$ and F^0 span the same column space, $V(r, F^0 H^r) = V(r, F^0)$. Thus the second item can be rewritten as $V(k, F^0 H^k) - V(r, F^0)$, which has a positive limit by Lemma 4. Hence $P[PC(k) < PC(r)] \rightarrow 0$ if $(r - k)g(N, T) \rightarrow 0$ as $N, T, k, r \rightarrow \infty$.

Next, for $k \geq r$,

$$P[PC(k) - PC(r) < 0] = P[V(r, \hat{F}^r) - V(k, \hat{F}^k) > (k - r)g(N, T)].$$

By Lemma 5, $V(r, \hat{F}^r) - V(k, \hat{F}^k) = O_p\left(\max\left\{\frac{k_{\max}^3}{\sqrt{N}}, \frac{k_{\max}^{2.5}}{\sqrt{T}}\right\}\right)$. According to our setting, $(k - r)g(N, T)$ converges to zero at a slower rate than $O_p\left(\max\left\{\frac{k_{\max}^3}{\sqrt{N}}, \frac{k_{\max}^{2.5}}{\sqrt{T}}\right\}\right)$. Thus, for $k > r$, $P[PC(k) < PC(r)] \rightarrow 0$ as $N, T, k, r \rightarrow \infty$. \square

Appendix B

See Figs. 1 and 2.

References

- Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203–1127.
- Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate factor models. *Statist. Probab. Lett.* 80, 1806–1813.
- Bai, C., Li, Q., Ouyang, M., 2014. Property taxes and home prices: a tale of two cities. *J. Econometrics* 180, 1–15.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Ching, H.S., Hsiao, C., Wan, S.K., 2012. Impact of CEPA on the labor market of Hong Kong. *China Econ. Rev.* 23, 975–981.
- Connor, G., Korajczyk, R., 1993. A test for the number of factors in an approximate factor model. *J. Finance* 48, 1263–1291.
- Cragg, J., Donald, S., 1997. Inferring the rank of a matrix. *J. Econometrics* 76, 223–250.
- Donald, S., 1997. Inference concerning the number of factors in a multivariate nonparametric relationship. *Econometrica* 65, 103–132.
- Fan, J., Liao, Y., Mincheva, M., 2011. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* 39, 3320–3356.
- Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 75, 603–680.
- Forni, M., Gambetti, L., 2010. Macroeconomic shocks and the business cycle: evidence from a structural factor model fiscal. Working Papers.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., (2000). Reference cycles: the NBER methodology revisited.
- Forni, M., Reichlin, L., 1998. Let's get real: a factor-analytic approach to disaggregated business cycle dynamics. *Rev. Econom. Stud.* 65, 453–473.
- Giannone, D., Reichlin, L., Sala, L., 2005. Monetary policy in real time. In: Gertler, M., Rogoff, K. (Eds.), *NBER Macroeconomics Annual 2004*. pp. 161–200.
- Gregory, A., Head, A., 1999. Common and country-specific fluctuations in productivity, investment, and the current account. *J. Monetary Econ.* 44, 423–452.
- Hallin, M., Liska, R., 2007. Determining the number of factors in the general dynamic factor model. *J. Amer. Statist. Assoc.* 102, 603–617.
- Hsiao, C., Ching, H.S., Wan, S.K., 2012. A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with Mainland China. *J. Appl. Econometrics* 27, 705–740.
- Jurado, K., Ludvigson, S., Ng, S., 2015. Measuring uncertainty. *American Economic Review* 105, 1177–1216.
- Lewbel, A., 1991. The rank of demand systems: theory and nonparametric estimation. *Econometrica* 59, 711–730.
- Li, K., Bell, D.R., 2017. Estimation of average treatment effects with panel data: theory and implementation. *J. Econometrics* (forthcoming).
- Ludvigson, S., Ng, S., 2007. The empirical risk-return relation: a factor analysis approach. *J. Financ. Econ.* 83, 171–222.
- Ludvigson, S., Ng, S., 2009. Macro factors in bond risk premia. *Rev. Financ. Stud.* 22, 5027–5067.
- Ludvigson, S., Ng, S., 2010. A factor analysis of bond risk premia. In: Uhla, A., Giles, D.E.A. (Eds.), *Handbook of Empirical Economics and Finance*. pp. 313–372.
- Mallows, C.L., 1973. Some comments on C_p . *Technometrics* 15, 661–675.
- Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447–1479.
- Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* 92, 1004–1016.
- Ouyang, M., Peng, Y., 2015. The treatment-effect estimation: A case study of the 2008 economic stimulus package of China. *J. Econometrics* 188, 545–557.
- Ross, S., 1976. The arbitrage theory of capital asset pricing. *J. Finance* 13, 341–360.
- Stock, J.H., Watson, M., (1998). Diffusion indexes. NBER Working Paper.
- Stock, J.H., Watson, M., 1999. Forecasting inflation. *J. Monetary Econ.* 44, 293–335.