Determining the Number of Factors When the Number of Factors Can Increase with Sample Size

Hongjun Li
International School of Economics and Management,
Capital University of Economics and Business, Beijing, 100070, PR China

Qi Li*
Department of Economics, Texas A&M University
and
International School of Economics and Management,
Capital University of Economics and Business, Beijing, 100070, PR China

Yutang Shi
Department of Economics, Texas A&M University
College Station, TX 77843-4228

December 15, 2013

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 criterion 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.

JEL classication: C2, C3, C5, G1

Keywords: Principal Components, Factor Analysis, Increasing Number of Factors, Information Criterion.

*We would like to thank seminar participants at Princeton University and Texas A&M University for useful comments that help to improve our paper. The corresponding author: Qi Li. Email: qi@econmail.tamu.edu.

1 Introduction

Factor models have been widely used in economic analyses such as forecasting economic variables, estimating variance-covariance matrix with high dimension data, and estimating average treatment effects. 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, 1999) consider forecasting inflation with diffusion indices ("factors") constructed from a large number of macroeconomic series. Gregory and Head (1999) and Forni, Hallin, Lippi, and Reichlin (2000) find that cross country variations have common components. Fan, Liao and Mincheva (2011), and Fan, Liao and Mincheva (2013) use factors model to estimate high dimensional variance-covariance matrix. Factor models can be used to evaluate the impacts of various polices. By assuming that the cross-sectional correlations for all the units are attributed to the presence of some (unobserved) common factors, Hsiao, Ching and Wan (2012) offer a panel data method to construct the counterfactuals and to measure average treatment effects of 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), Anh 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, Reichlin and Sala (2005) and Forni and Gambetti (2010). This suggests that the number of factors may dependent on sample sizes. One reason that the number of factors may increase with sample size is structure break, new factors may emerge after economic environments change. Using Bai and Ng's (2002) 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, Ludvigson and Ng (2013) 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, Ludvigson and Ng's (2013) finding supports the argument that the number of factors may increase as sample 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 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. Onatski (2009) develops a test of the null of k_0 factors against the alternative that the number of factors r is $k_0 < r \leq k_1$ 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. 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 $N, T \to \infty$. 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 estimating 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 ($\widehat{F}^k$) using 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 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 $(\widehat{F}^k)$. 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.

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 $r = r_{N,T} \to \infty$, as $N \to \infty$, or $T \to \infty$, or both $N, T \to \infty$, but with a slower rate than $\min\{N, T\}$, i.e., $\max\{r/N, r/T\} \to 0$, as $N, T \to \infty$.

Let X_{it} denote the response variable for unit i at time t, for $i = 1, \ldots, N$, and $t = 1, \ldots, T$. The factor structure is of the form

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

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 component of the response variable X_{it}. There are no deterministic terms. The factors, factor loadings and idiosyncratic components are not observed. The matrix form of the factor model is

$$X = F^0 \Lambda^{0'} + e, \tag{2}$$

where X is a $T \times N$ matrix $(X_{ti})_{(T \times N)}$, $F^0 = (F_1, F_2, \ldots, F_T)'$ is the $T \times r$ matrix of factors, $\Lambda^0 = (\lambda_1^0, \lambda_2^0, \ldots, \lambda_N^0)'$ is the $N \times r$ matrix of factor loadings, and $e = (e_{it})_{(T \times N)}$ is the $T \times N$ matrix of idiosyncratic components.

Let $\mathrm{tr}(A)$ denote the trace of a square matrix A. The norm of matrix A is defined as $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$. Let m denote the minimum of N and T. M and C denote some generic positive constants. $\mathcal{N}$ denotes the set of natural number. We make the main assumptions as follows:

ASSUMPTION A (Factors): $\sup_{r\in\mathcal{N}} r^{-2}E\|F_t^0\|^4 < M$. Also, 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 $T \to \infty$.

ASSUMPTION B (Factors Loadings): $\max_{1\le i\le N} r^{-2}E\|\lambda_i^0\|^4 \le C < \infty$, and there exists a $r \times r$ positive definite matrix D such that $\|\Lambda^{0'}\Lambda^0/N - D\| \xrightarrow{p} 0$ as $N \to \infty$.

ASSUMPTION C (Idiosyncratic Components): As $N, T \to \infty$,

1. $E(e_{it}) = 0, E|e_{it}|^8 \le M$;
2. $E(e_s' e_t) = E(N^{-1}\sum_{i=1}^N e_{is}e_{it}) = \gamma_N(s,t),\ |\gamma_N(s,s)| \le M$ for all s, and $T^{-1}\sum_{s=1}^T\sum_{t=1}^T |\gamma_N(s,t)| \le M$;
3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \le |\tau_{ij}|$ for some τ_{ij} and for all t; furthermore, $N^{-1}\sum_{i=1}^N\sum_{j=1}^N |\tau_{ij}| \le M$;
4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T |\tau_{ij,ts}| \le M$;
5. for every (t,s), $E|N^{-1/2}\sum_{i=1}^N [e_{is}e_{it} - E(e_{is}e_{it})]|^4 \le M$.
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 7^{th} moment and $E(\varepsilon_{it}) = 0$.

ASSUMPTION D (Weak Dependence Between Factors and Idiosyncratic Components):

$$E\left(\frac{1}{N}\sum_{i=1}^N \left\|\frac{1}{\sqrt{Tr}}\sum_{t=1}^T F_t^0 e_{it}\right\|^2\right) \le M.$$

Assumptions A-B are modified from Assumptions A-B in Bai and Ng (2002). They are quite standard for factor models. Assumptions C-D are also similar to Assumptions C-D in Bai and Ng (2002), which allow for limited time-series and cross-section dependence in idiosyncratic component and also weak dependence between factors and idiosyncratic errors. Assumption C(6) puts a structure on the idiosyncratic components.

This structure allows heteroskedasticity in both the time and cross-section dimensions, and also limited autocorrelation and cross-sectional correlation in the components.

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 $k \in \{1, \ldots, k_{max}\}$, 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 the loadings and factors with the allowance of k factors in the estimation. The method of principal components minimizes

$$V(k) = \min_{\Lambda^k, F^k} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k'} F_t^k)^2, \tag{3}$$

subject to the normalization of either $\Lambda^{k'}\Lambda^k/N = I_k$ or $F^{k'}F^k/T = I_k$.

Let $\text{ev}_{(i)}(A)$ denote the i^{th} largest eigenvalue of matrix A, and $\text{EV}_{(i)}(A)$ is the eigenvector corresponding to the eigenvalue $\text{ev}_{(\text{i})}(A)$ of matrix A. If we concentrate out F^k and use the normalization that $\Lambda^{k'}\Lambda^k/N = I_k$, the solution to the above problem is given by $(\bar{F}^k, \bar{\Lambda}^k)$, where $\bar{\Lambda}^k = \sqrt{T}(\text{EV}_{(1)}(X'X), \ldots, \text{EV}_{(k)}(X'X))$. The normalization that $\Lambda^{k'}\Lambda^k/N = I_k$ implies $\bar{F}^k = X\bar{\Lambda}^k/N$. Define $\widehat{F}^k = \bar{F}^k(\bar{F}^{k'}\bar{F}^k)^{1/2}$, a rescaled estimator of the factors. This rescaled estimator has the asymptotic properties summarized in the following theorem.

THEOREM 3.1 *For any* $1 \leq k \leq k_{\max} = o(\min\{N^{1/3}, T\})$ *there exists a* $(r \times k)$ *matrix* H^k *with* $\text{rank} = \min\{k, r\}$ *such that*

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_t^k - H^{k'} F_t^0 \right\|^2 = O_p\left(\max\left\{ \frac{kr^2}{N}, \frac{k}{T} \right\} \right). \tag{4}$$

Similar to the results of Bai and Ng (2002), Theorem 3.1 suggests that the time average of the squared deviations between the estimated factors $\widehat{F}^k$ and those that lie in the true factor space, $H^{k'}F_t^0$, will vanish as $N, T \to \infty$. 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 Theorem 3.1, we can now analyze the problem of determining the number of factors. Let $V(k, F^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k'} F_t^k)^2$ be the sum of squared residuals (divided by NT) from time-series regressions of $\underline{X}_i$ on the k factors for all $i = 1, \ldots, N$. The selecting criteria modified from those suggested by Bai and Ng (2002) have the form

$$PC(k) = V(k, \widehat{F}^k) + kg(N,T), \tag{5}$$

where $g(N,T)$ is the penalty factor satisfying two conditions: (i) $k_{max} \cdot g(N,T) \to 0$ as $N,T \to \infty$, (ii) $C^{-1}_{N,T,k_{max}} g(N,T) \to \infty$ as $N,T \to \infty$,where $C_{N,T,k_{max}} = O_p\left(\max\left\{\frac{k^6_{max}}{N}, \frac{k^4_{max}}{T}\right\}\right)$. As $V(k, \widehat{F}^k)$ is decreasing in k, the criteria above penalize k with a penalty factor $g(N,T)$ to select the estimator $\hat{k}$ such that under and overparameterized models will not be chosen. Theorem 3.2 establishes this result formally.

THEOREM 3.2 *Let $1 \le r \le k_{max} = o(\min\{N^{1/16}, T^{1/14}\})$ and $\hat{k} = \text{argmin}_{1 \le k \le k_{max}} PC(k)$. Suppose that Assumptions A-D hold, and that (i) $k_{max} \cdot g(N,T) \to 0$, (ii) $C^{-1}_{N,T,k_{max}} \cdot g(N,T) \to \infty$ as $N,T \to \infty$. Then*

$$\lim_{N,T \to \infty} \text{Prob}[\hat{k} = r] = 1. \tag{6}$$

A formal proof of Theorem 3.2 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 the consistent estimation so that they may not always be required for consistent estimating the number of factors. Similar to Bai and Ng (2002), we also have the following result[1]:

Corollary 3.1: *Under the Assumptions of Theorem 3.2, if one replaces $PC(k)$ in Theorem 3.2 by the class of criterion defined by*

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

then the conclusion of Theorem 3.2 holds true.

Corollary 3.1 states that the class of criterion $PC(k)$ 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 $\widehat{\sigma}^2$ 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 the three PC_p criteria and as follows:

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

[1]The proof of this result is omitted as it is almost the same as the proof of Corollary 1 in Bai and Ng (2002).

It is easy to check that these criteria satisfy the two conditions for the penalty factor in Theorem 3.2 if $k_{max} = o_p\left(\left[\ln\left(\frac{NT}{N+T}\right)\right]^{1/6}\right)$. These 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 $V(k_{max}, \widehat{F}^{k_{max}}) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2$, where $\hat{e}_{it} = X_{it} - \widehat{\lambda}_i^{k_{max}}\widehat{F}_t^{k_{max}}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, the residuals for the linear regression of X on $\widehat{F}^{k_{max}}$. Thus, the number of factors estimated using these three criteria may be sensitive to the selection of k_{max}. Corollary 3.1 suggests the following three IC_p criteria can also be used to select the number of factors:

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

The main advantage of these 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 selecting k_{max} using this rule can lead to $k_{max} < r$ since r increases with N or T in our case, which will lead to an underestimation of the number of factors because $\hat{k} \leq k_{max} < r$. On the other hand, if k_{max} is too large ($k_{max} >> r$), the selected $\hat{k}$ tend 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. We use a specific example to illustrate this selection procedure. We generate a simulated data of $N = 100$, $T = 60$ with the true number of factors $r = 7$. We let k_{max} take values from $\{1, 2, ..., 40\}$. 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 Figure 1. From Figure 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

$k_{max} = 7, 8, ..., 16$), 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).

Figure 2 plot $\hat{k}$-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 Figure 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} to take values in $\{1, 2, ..., 40\}$ since $r \leq 40$ is likely to be true for the panel data sets economists encounter in practice.

4 Simulations

In this section we conduct Monte Carlo simulation 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 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 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 form as follows:

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

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

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

We consider four DGPs here. In the base case, we set the DGP as $\theta = 1$ and $e_{it} \sim i.i.d.N(0,1)$. This base DGP is denoted as DGP1. The simulation results for this case are reported in Table 1 where the boldfaced numbers indicate incorrect selection of the number of factors. We see that PC_{p3} selects a $\hat{k}$ larger than r in some cases. When the sample sizes are large, i.e. $\min\{N,T\} > 60$, PC_{p1}, PC_{p2}, IC_{p1} and IC_{p2} give precise estimates of the number of factors.

Table 1: Estimated Number of Factors: DGP1

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	**12**	6	**8**	6
200	60	7	7	7	**8**	7	**8**	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	**7**	**6**	**11**	5	5	**7**
60	100	6	6	6	**12**	6	6	**9**
100	100	6	6	6	**13**	6	6	**11**
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	**6**	**6**	**10**	5	5	5
100	60	6	6	6	**12**	6	6	6
60	100	6	6	6	**12**	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: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; \theta = 1$, $r = [c * \ln(N)]$ *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*.*

We denote the high-variance case as DGP2, let $\theta = 5$ and keep all the other parameters the same as those of DGP1. The estimated results of $\hat{k}$ are reported in Table 2. It is similar to the base case that PC_{p1}, PC_{p2} and IC_{p3} perform pretty well and give precise estimates of the number of factors for almost all cases.

For the heterogeneity case of DGP3, we set the idiosyncratic shocks to be heterogeneous. We let $\theta = 5$, and $e_{it} = u_{it} + \delta_t \epsilon_{it}$ 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.

Table 2: Estimated number of Factors: High-Variance

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	**12**	6	6	6
200	60	7	7	7	**8**	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	**7**	5.002	**10.002**	5	5	5
60	100	6	6	6	**12**	6	6	6
100	100	6	6	6	**13**	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	10	11	11	11	11	11	11
100	40	5	5	**4**	5	**1**	**0**	5
100	60	6	**5**	**5**	6	**1**	**1**	**5**
60	100	6	6	6	**12**	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: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; \theta = 5$, $r = [c \ln(N)]$ *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.*

Thus the variance of the idiosyncratic shocks is 5 when t is odd and 10 when t is even. We denote this DGP as DGP3. The estimated values of $\hat{k}$ are reported in Table 3. Similar to the homogeneous cases, PC_{p1} and PC_{p2} perform well when the sample sizes are large. But IC_{P1} and IC_{p2} tend to select $\hat{k}$ that is smaller than the true number of factors r, while PC_{p3} tends to overestimate r.

For the last case, denoted as DGP4, we allow the idiosyncratic to be autocorrelated. We set $\theta = 5$ and $e_{it} = \rho * e_{it-1} + v_{it}$, where $\rho = 0.5$ and $v_{it} \sim i.i.d.N(0, 1)$. The estimating results are reported in Table 4. The results for this case are almost the same as those of the base case. When the sample sizes are large, i.e. $\min\{N, T\} > 40$, PC_{p1}, PC_{p2}, IC_{p1} and IC_{p2} perform quite well in accurately estimating the number of factors.

Summarizing the results for all the DGPs we observe that PC_{p1} and PC_{p2} have the best overall perfor-

Table 3: Estimated number of Factors: Heterogeneity

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6.002	6	**11.002**	**3**	**1**	**6**
200	60	7	7	7	**8**	**4**	**3**	7
500	60	9	9	9	9	**8**	**8**	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	10	10	10
40	100	5	**7**	**6**	**11**	**1**	**1**	**4**
60	100	6	6	6	**13**	**3**	**1**	**6**
100	100	6	6	6	**11**	**5**	**4**	6
200	100	7	7	7	7	**6**	**6**	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	**7**	5	**10**	**1**	**0**	**5**
100	60	6	6	6	**11**	**1**	**1**	**5**
60	100	6	6	6	**13**	**3**	**1**	**6**
60	200	7	7	7	**8**	**4**	**3**	**7**
60	500	9	9	9	9	**8**	**8**	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	10	10	10

DGP3: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; e_{it} = u_{it} + \delta_t \epsilon_{it}$, *where* $\delta_t = 0$ *for* t *even, and* $\delta_t = 1$ *for* t *odd;* $\theta = 5$, $r = [c \ln(N)]$ *for the upper part of the table (divided with), and* $r = [c \ln(T)]$ *for the lower part, where* $[A]$ *denotes taking the integer part of a real number.*

mance. Hence, we recommend using the PC_{p1} and PC_{p2} criteria in practice.

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, i.e. $r = o(\min\{N^{1/16}, T^{1/14}\})$. We extend the analysis of Bai and Ng (2002) to the case that number of factors can increase and prove the consistency of Bai and Ng's (2002) procedure in determining the number of factors. We also propose a new procedure so that our selected number of factors is not sensitive to the selection of k_{max}. Monte Carlo simulation results suggest that the criteria PC_{p1}, PC_{p2} have the overall best performance and therefore can be used to accurately estimate the number of factors when the data

Table 4: Estimated number of Factors: Autocorrelation

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	**12**	6	6	6
200	60	7	7	7	**8**	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	**7**	**6**	**10**	5	**3**	5
60	100	6	6	6	**12**	6	6	**7**
100	100	6	6	6	**13**	6	6	**11**
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	9	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	**7**	5	**11**	5	5	5
100	60	6	6	6	**12**	6	6	6
60	100	6	6	6	**12**	6	6	6
60	200	7	7	7	**8**	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

DGP4: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}; e_{it} = \rho e_{it-1} + v_{it}$; $\rho = 0.5, \theta = 5$; $r = [c \ln(N)]$ *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.*

dimensions are relatively large, say $\min\{N, T\} \geq 60$.

One possible future research topic is to find alternative criteria that can improve the finite-sample performance 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.

Appendix A: Proofs

Proof of Theorem 3.1

We will first prove a lemma (Lemma 1) below which will be used to Theorem 3.1.

Lemma 1 Under Assumptions A-C, we have for some $M_1 < \infty$, and for all N and T,

(i) $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,t)^2 \leq M_1$;

(ii) $E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|(Nr)^{-\frac{1}{2}}e_t'\Lambda^0\right\|^2\right) = E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|(Nr)^{-\frac{1}{2}}\sum_{i=1}^{N}e_{it}\lambda_i^0\right\|^2\right) \leq M_1$;

(iii) $E\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right) \leq M_1$;

(iv) $E\left\|(NTr)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\lambda_i^0\right\| \leq M_1$

Proof :

(i) Same as Lemma 1(i) in Bai and Ng (2002).

(ii)

$$
\begin{aligned}
& E\left(\left\|(Nr)^{-\frac{1}{2}}\sum_{i=1}^{N}e_{it}\lambda_i^0\right\|^2\right) \\
=& \frac{1}{Nr}\sum_{i=1}^{N}\sum_{j=1}^{N}E(e_{it}e_{jt})E(\lambda_i^{0\prime}\lambda_j^0) \\
=& \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\tau_{ij,t}E\left(\frac{\lambda_i^{0\prime}\lambda_j^0}{r}\right) \\
\leq& CM
\end{aligned}
$$

by Assumptions B and C(3).

(iii) Same as Lemma 1(iii) in Bai and Ng (2002).

(iv)

$$
\begin{aligned}
& E\left\|(NTr)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}\lambda_i^0\right\|^2 \\
=& \frac{1}{NTr}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E(e_{it}e_{js})E\left(\lambda_i^{0'}\lambda_j^0\right) \\
\leq & C\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \\
\leq & CM
\end{aligned}
$$

by Assumptions B and C(4). ■

Proof of Theorem 3.1:

Recall that $\hat{F}^k = N^{-1}X\tilde{\Lambda}^k$ and $\tilde{\Lambda}^k = T^{-1}X'\tilde{F}^k$. 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), $H^{k'} = (\tilde{F}^{k'}F^0/T)(\Lambda^{0'}\Lambda^0/N)$, we have

$$
\hat{F}_t^k - H^{k'}F_t^0 = \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^k\xi_{st}
$$

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

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 $a_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\gamma_N(s,t)\right\|^2$, $b_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st}\right\|^2$, $c_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\eta_{st}\right\|^2$ and $d_t = \frac{1}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\xi_{st}\right\|^2$. 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 $\|\sum_{s=1}^{T}\tilde{F}_s^k\gamma_N(s,t)\|^2 \leq \left(\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\cdot\left(\sum_{s=1}^{T}\gamma_N(s,t)^2\right)$. Thus,

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

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

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

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 Bai and Ng (2002).

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

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

by Assumption C5, we have

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

For c_t, we have

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

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

It follows that

$$\frac{1}{T}\sum_{t=1}^{T} c_t = O_p(kr)\frac{r}{N}\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{e_t'\Lambda^0}{\sqrt{Nr}}\right\|^2 = O_p\left(\frac{kr^2}{N}\right)$$

because $\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{e_t'\Lambda^0}{\sqrt{Nr}}\right\|^2 = O_p(1)$ by Lemma 1(ii).

The term $(1/T)\sum_{t=1}^{T} d_t = O_p\left(\frac{kr^2}{N}\right)$ can be proved similarly. Combining the above results, we have shown that

$$\begin{aligned}(1/T)\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2 &\leq (4/T)\sum_{t=1}^{T}(a_t + b_t + c_t + d_t) \\ &= O_p\left(\frac{r^2 k}{N}\right) + O_p\left(\frac{k}{T}\right).\end{aligned}$$

■

Proof of Theorem 3.2

Lemma 2 Let $D_k = \frac{\hat{F}^{k'}\hat{F}^k}{T}$ and $D_0 = \frac{H^{k'}F^{0'}F^0H^k}{T}$. When $k \leq r$, we have (i) $\|D_k^{-1}\| = O_p(k)$; (ii) $\|D_k^{-1} - D_0^{-1}\| = O_p\left(\max\left\{\frac{r^2k_{max}^3}{\sqrt{N}}, \frac{rk_{max}^3}{\sqrt{T}}\right\}\right)$.

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

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

Hence, we have

$$\begin{aligned}\|D_k - D_0\| &\leq \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2 + 2\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2\right)^{1/2} \\ &= O_p\left(\max\left\{\frac{r^2k}{N}, \frac{k}{T}\right\}\right) + O_p\left(\max\left\{\frac{r\sqrt{k}}{\sqrt{N}}, \frac{\sqrt{k}}{\sqrt{T}}\right\}\right)\cdot O_p\left(r\sqrt{k}\right) \\ &= O_p\left(\max\left\{\frac{r^2k}{\sqrt{N}}, \frac{rk}{\sqrt{T}}\right\}\right)\end{aligned}$$

by Theorem 3.1 and the fact that $\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2 = O_p(r^2 k_{max})$, which is shown below.

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

$$\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2 - E\left[\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2\right] = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Since

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2\right] &= E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{k}\left(\sum_{j=1}^{r}H_{lj}^{k'}F_{tj}^0\right)^2\right] \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{k}\sum_{j=1}^{r}\sum_{i=1}^{r}E\left[H_{lj}^{k'}F_{tj}^0 H_{li}^{k'}F_{ti}^0\right] \\ &= O_p(r^2 k_{max}), \end{aligned}$$

where the last equality uses the fact that $E\left[H_{lj}^{k'}F_{tj}^0 H_{li}^{k'}F_{ti}^0\right]$ is finite for all $t = 1, \ldots, T$, and $l, j, i = 1, \ldots, r$. Thus, we have shown that $\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|^2 = O_p(r^2 k_{max})$.

Since $\widehat{F}^k = \bar{F}^k(\bar{F}^{k'}\bar{F}^k)^{(1/2)}$ and $\bar{F}^k = X\bar{\Lambda}^k/N$,

$$\begin{aligned} \frac{\widehat{F}^{k'}\widehat{F}^k}{T} &= \frac{1}{T}(\bar{F}^{k'}\bar{F}^k)^{(1/2)}\bar{F}^{k'}\bar{F}^k(\bar{F}^{k'}\bar{F}^k)^{(1/2)} \\ &= \frac{1}{T}(\bar{F}^{k'}\bar{F}^k)^2 \\ &= \frac{1}{TN^4}\left(\bar{\Lambda}^{k'}X'X\bar{\Lambda}^k\right)^2 \\ &= \frac{1}{TN^4}\left(\bar{\Lambda}^{k'}\bar{\Lambda}^k D_{EV}\right)^2 \\ &= \left(\frac{D_{EV}}{N}\right)^2, \end{aligned}$$

where D_{EV} is the $k \times k$ diagonal matrix consisting of $(ev_{(1)}(X'X), \ldots, ev_{(k)}(X'X))$. Bai and Ng (2008) have a similar identity. This implies $k^{-1}\|D_k^{-1}\| = O_p(1)$. Similarly, given $H^k = \frac{\Lambda^{0'}\Lambda^0}{N}\frac{F^{0'}\tilde{F}^k}{T}$, we have that $k^{-1}\|D_0^{-1}\| = O_p(1)$.

From $D_k^{-1} - D_0^{-1} = D_k^{-1}(D_0 - D_k)D_0^{-1}$, we have

$$\begin{aligned}
\|D_k^{-1} - D_0^{-1}\| &= \|D_k^{-1}(D_o - D_k)D_0^{-1}\| \\
&\leq \|D_k^{-1}\| \cdot \|D_0 - D_k\| \cdot \|D_0^{-1}\| \\
&= k^2 \frac{\|D_k^{-1}\|}{k} \cdot \|D_0 - D_k\| \cdot \frac{\|D_0^{-1}\|}{k} \\
&= k^2 \cdot O_p(1) \cdot O_p\left(\max\left\{\frac{kr^2}{\sqrt{N}}, \frac{kr}{\sqrt{T}}\right\}\right) \\
&= O_p\left(\max\left\{\frac{r^2 k_{max}^3}{\sqrt{N}}, \frac{r k_{max}^3}{\sqrt{T}}\right\}\right).
\end{aligned}$$

■

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

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = O_p\left(\max\left\{\frac{r^4 k_{max}^4}{\sqrt{N}}, \frac{r^3 k_{max}^4}{\sqrt{T}}\right\}\right).$$

Proof :

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

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

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

$$\begin{aligned}
P_{\hat{F}}^k - P_{FH}^0 &= \frac{1}{T}\hat{F}^k\left(\frac{\hat{F}^{k'}\hat{F}^k}{T}\right)^{-1}\hat{F}^{k'} - \frac{1}{T}F^0 H^k\left(\frac{H^{k'}F^{0'}F^0H^k}{T}\right)^{-1} H^{k'}F^{0'} \\
&= \frac{1}{T}[\hat{F}^{k'} D_k^{-1}\hat{F}^k - F^0 H^k D_0^{-1} H^{k'} F^{0'}] \\
&= \frac{1}{T}\left[(\hat{F}^k - F^0H^k + F^0H^k)D_k^{-1}(\hat{F}^k - F^0H^k + F^0H^k)' - F^0H^kD_0^{-1}H^{k'}F^{0'}\right] \\
&= \frac{1}{T}[(\hat{F}^k - F^0H^k)D_k^{-1}(\hat{F}^k - F^0H^k)' + (\hat{F}^k - F^0H^k)D_k^{-1}H^{k'}F^{0'} \\
&\quad + F^0H^kD_k^{-1}(\hat{F}^k - F^0H^k)' - F^0H^kD_0^{-1}H^{k'}F^{0'}].
\end{aligned}$$

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.

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

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

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

It can be verified that III is also $O_p\left(\max\left\{\frac{k^2r^2}{\sqrt{N}}, \frac{k^2r}{\sqrt{T}}\right\}\right)$.

$$
\begin{aligned}
IV &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t^{0'}H^k(D_k^{-1} - D_0^{-1})H^{k'}F_s^0X_{it}X_{is} \\
&\leq \|D_k^{-1} - D_0^{-1}\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\|\cdot|X_{it}|\right)^2 \\
&\leq \|D_k^{-1} - D_0^{-1}\|\frac{kr^2}{N}\sum_{i=1}^{N}\left(\frac{1}{T\sqrt{kr}}\sum_{t=1}^{T}\|H^{k'}F_t^0\|\right)^2 \\
&= \|D_k^{-1} - D_0^{-1}\|\cdot kr^2\cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^4r^4}{\sqrt{N}}, \frac{k^4r^3}{\sqrt{T}}\right\}\right).
\end{aligned}
$$

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

Thus, we have

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = O_p\left(\max\left\{\frac{r^4 k_{max}^4}{\sqrt{N}}, \frac{r^3 k_{max}^4}{\sqrt{T}}\right\}\right).$$

■

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

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

Proof :

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

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

$$\begin{aligned}
I &= \frac{1}{NT}\sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_{FH}^0 F^0 \lambda_i^0 \\
&= \frac{1}{NT}\sum_{i=1}^{N} \left(M_{FH}^0 F^0 \lambda_i^0\right)' M_{FH}^0 F^0 \lambda_i^0 \\
&\geq C > 0
\end{aligned}$$

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

Next,

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

Consider the first term

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

where the last equality follows from Lemma 1(ii). The second term is also $o_p(1)$, and hence $II = o_p(1)$. ■

Lemma 5 For any k with $r \leq k \leq k_{max}$, $V(k,\hat{F}^k) - V(r,\hat{F}^r) = O_p\left(\max\left\{\frac{r^4k_{max}^2}{N}, \frac{r^2k_{max}}{T}\right\}\right)$.

Proof :

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

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

$$
V(k,\hat{F}^k) - V(r,F^0) = O_p\left(\max\left\{\frac{r^4k_{max}^2}{N}, \frac{r^2k_{max}}{\sqrt{T}}\right\}\right).
$$

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

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

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

Note that

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

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

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

by Theorem 3.1.

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

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

by Theorem 3.1 and Lemma 1(ii). Therefore,

$$V(k, \hat{F}^k) = \frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' M_{\hat{F}}^k \underline{e}_i + O_p \left(\max \left\{ \frac{k^2 r^4}{N}, \frac{k^2 r^2}{T} \right\} \right)$$

Thus we have

$$V(k, \hat{F}^k) - V(r, F^0) = \frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' P_F^0 \underline{e}_i - \frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' P_{\hat{F}}^k \underline{e}_i + O_p \left(\max \left\{ \frac{k^2 r^4}{N}, \frac{k^2 r^2}{T} \right\} \right)$$

Note that

$$\begin{aligned}
\frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' P_F^0 \underline{e}_i &\leq \left\| \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right\| \cdot \frac{1}{NT^2} \sum_{i=1}^{N} \underline{e}_i' F^0 F^{0\prime} \underline{e}_i \\
&= \left\| \left(\frac{F^{0\prime} F^0}{T} \right)^{-1} \right\| \cdot \frac{1}{NT} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{Tr}} \sum_{t=1}^{T} F_t^0 \underline{e}_{it} \right\|^2 \cdot r \\
&= r \cdot O_p(1) \cdot \frac{1}{T} \cdot r \cdot O_p(1) \\
&= O_p \left(\frac{r^2}{T} \right) \leq O_p \left(\max \left\{ \frac{k^2 r^4}{N}, \frac{k^2 r^2}{T} \right\} \right)
\end{aligned}$$

$\frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' P_F^0 \underline{e}_i$ is bounded by the sum of the first k largest eigenvalues of the matrix $A_{NT} = \frac{1}{NT} e'e$, where $e = (e_{ti}), T \times N$. Let $\rho(A)$ denote the largest eigenvalue of a matrix A. Under Assumption C(6), as Bai and Ng (2005) shows, $\rho(A_{NT}) = O_p(C_{NT}^{-2})$, where $C_{NT}^2 = \min(N, T)$. Thus,

$$\frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' P_F^0 \underline{e}_i = O_p \left(\max \left\{ \frac{k}{N}, \frac{k}{T} \right\} \right) \leq O_p \left(\max \left\{ \frac{k^2 r^4}{N}, \frac{k^2 r^2}{T} \right\} \right)$$

In summary,

$$V(k, \hat{F}^k) - V(r, F^0) = O_p \left(\max \left\{ \frac{r^4 k_{max}^2}{N}, \frac{r^2 k_{max}}{T} \right\} \right)$$

■

Proof of Theorem 3.2

Proof :

We shall prove that $\lim_{N,T \to \infty} P(PC(k) < PC(r)) = 0$ 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 $P[V(k,\hat{F}^k) - V(r,\hat{F}^r) < (r-k)g(N,T)] \to 0$ as $N,T,k,r \to \infty$.

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

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

Lemma 2 implies that the first and the third terms are both $O_p\left(\max\left\{\frac{k_{max}^8}{\sqrt{N}}, \frac{k_{max}^7}{\sqrt{T}}\right\}\right)$. Next, we consider the second item. Because F^0H^r and F^0 span the same column space, $V(r,F^0H^r) = V(r,F^0)$. Thus the second item can be rewritten as $V(k,F^0H^k) - V(r,F^0)$, which has a positive limit by Lemma 3. Hence $P[PC(k) < PC(r)] \to 0$ if $(r-k)g(N,T) \to 0$ as $N,T,k,r \to \infty$.

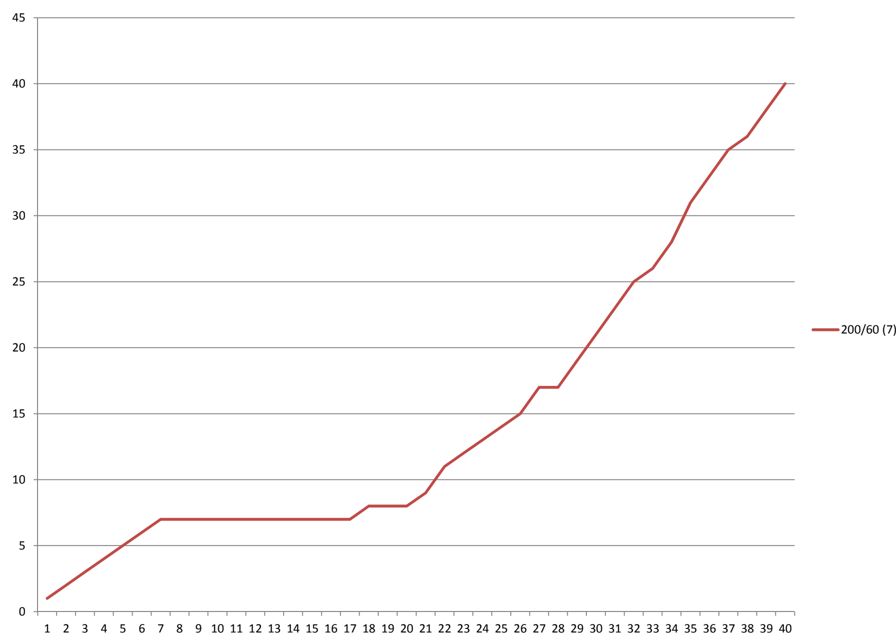
Next, for $k \geq r$,

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

By Lemma 4, $V(r,\hat{F}^r) - V(k,\hat{F}^k) = O_p\left(\max\left\{\frac{k_{max}^6}{N}, \frac{k_{max}^4}{T}\right\}\right)$. According to our setting, $(k-r)g(N,T)$ converges to zero at a slower rate than $O_p\left(\max\left\{\frac{k_{max}^6}{N}, \frac{k_{max}^4}{T}\right\}\right)$. Thus, for $k > r$, $P[PC(k) < PC(r)] \to 0$ as $N,T,k,r \to \infty$. ■

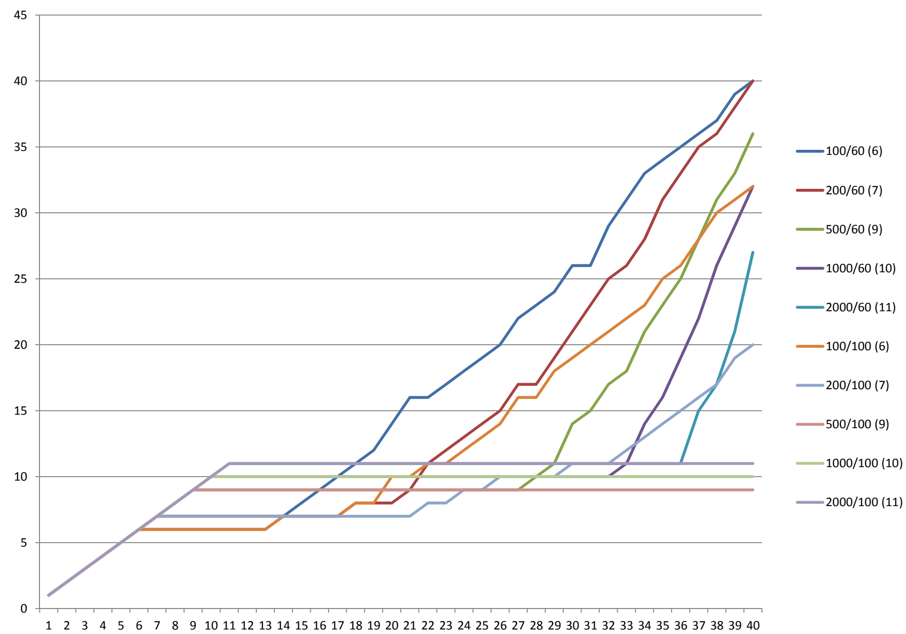
Appendix B: Figures

Figure 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 $k_{max} \in [1, 40]$.

Figure 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 $r = 6$.

References

[1] Anh, S. and A. Horenstein (2013), Eigenvalue Ratio Test for the Number of Factors, *Econometrica*, 81, 1203-1127.

[2] Bai, J. and S. Ng (2002), Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191-221.

[3] Connor, G., and R. Korajzcyk (1993), A Test for the Number of Factors in an Approximate Factor Model, *Journal of Finance*, 48, 1263-1291.

[4] Cragg, J., and S. Donald (1997), Inferring the Rank of a Matrix, *Journal of Econometrics*, 76, 223-250.

[5] Donald, S. (1997), Inference Concerning the Number of Factors in a Multivariate Nonparameteric Relationship, *Econometrica*, 65, 103-132.

[6] Fan, J., Y. Liao and M. Mincheva (2011), High-Dimensional Covariance Matrix Estimation in Approximate Factor Models, *Annals of Statistics*, 39, 3320-3356.

[7] Fan, J., Y. Liao and M. Mincheva (2013), Large Covariance Estimation by Thresholding Principal Orthogonal Complements, *Journal of the Royal Statistical Society: Series B*, 75, 603-680.

[8] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), Reference Cycles: The NBER Methodology Revisited, *CEPR Discussion Paper 2400.*

[9] Forni, M., and L. Gambetti, (2010), Macroeconomic Shocks and the Business Cycle: Evidence from a Structural Factor Model fiscal, *Working Papers 440*, Barcelona Graduate School of Economics.

[10] Forni, M., and L. Reichlin (1998), Lets Get Real: a Factor-Analytic Approach to Disaggregated Business Cycle Dynamics, *Review of Economic Studies*, 65, 453-473.

[11] Giannone, D., L. Reichlin and L. Sala (2005), Monetary Policy in Real Time, *NBER Macroeconomics Annual 2004*, e.d. by Mark Gertler and Kenneth Rogoff, 161-200. MIT Press.

[12] Gregory, A., and A. Head (1999), Common and Country-Specific Fluctuations in Productivity, Investment, and the Current Account, *Journal of Monetary Economics*, 44, 423-452.

[13] Hsiao, C., H. S. Ching and S. K. Wan (2012), A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China *Journal of Applied Econometrics*, 27, 705-740.

[14] Jurado, K., S. Ludvigson and S. Ng (2013), Measuring Uncertainty, *working paper.*

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

[16] Ludvigson, S. and S. Ng (2007), The empirical risk-return relation: a factor analysis approach, *Journal of Financial Economics*, 83, 171-222.

[17] Ludvigson, S. and S. Ng (2009), Macro Factors in Bond Risk Premia, *The Review of Financial Studies*, 22, 5027-5067.

[18] Ludvigson, S. and S. Ng (2010), A Factor Analysis of Bond Risk Premia, *Handbook of Empirical Economics and Finance*, e.d. by Aman Uhla and David E. A. Giles, 313-372. Chapman and Hall, Boca Raton, FL.

[19] Onatski, A. (2009), Testing Hypotheses about the Number of Factors in Large Factor Models, *Econometrica*, 77, 1447-1479.

[20] Onatski, A. (2010), Determing the Number of Factors from Empirical Distribution of Eigenvalues, *Review of Economics and Statistics*, 92, 1004-1016.

[21] Ross, S. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Finance*, 13, 341-360.

[22] Stock, J. H., and M. Watson (1998), Diffusion Indexes, *NBER Working Paper 6702.*

[23] Stock, J. H., and M. Watson (1999), Forecasting Inflation, *Journal of Monetary Economics*, 44, 293-335.