Jinshan Liu[a,b] / Jianzhu Pan[c] / Qiang Xia[d1] / Li Xiao[d]

On Determination of the Number of Factors in an Approximate Factor Model

[a]School of Financial Mathematics and Statistics, Guangdong University of Finance, Guangzhou 510521, China
[b] Key Laboratory of Advanced Theory and Application in Statistics and Data Science, East China Normal University, Ministry of Education
[c]Department of Mathematics & Statistics, University of Strathclyde, Glasgow G1 1XH, UK
[d]College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

Abstract:
This paper proposes a ridge-type method for determining the number of factors in an approximate factor model. The new estimator of factor number is obtained by maximizing both the ratio of two adjacent eigenvalues and the cumulative contribution rate of the factors which represents the explanatory power of the common factors for response variables. Our estimator is proved to be as asymptotically consistent as those in Ahn and Horenstein (2013). But Monte Carlo simulation experiments show our method has better correct selection rates in finite sample cases. A real data example is given for illustration.
Key words: Approximate factor model, number of factors, eigenvalue ratio, cumulative contribution rate, ridge-type method.
JEL classification: C1-C3

1 Introduction

One of effective approaches to dimensional reduction is factor analysis. A small number of common factors are found and used to represent a high dimensional data set. Many references on various factor models for economic or financial data with large numbers of cross-sectional units and a large sample of time series observations have appeared in literature. These factor models could be classified into three categories: Static approximate factor models proposed by Chamberlain and Rothschild (2014) and followed by Connor and Korajzcyk (1993), Bai and Ng (2002), Stock and Watson (2002), Onatski (2010), Alessi, Barigozzi, and Capasso (2010), Fan, Liao, and Mincheva (2013), Ahn and Horenstein (2013), Caner and Han (2014),

[1]Qiang Xia is the corresponding author. E-mail: xiaqiang@scau.edu.cn

This is a peer-reviewed, accepted author manuscript of the following article: Liu, J., Pan, J., Xia, Q., & Xiao, L. (Accepted/In press). On determination of the number of factors in an approximate factor model. Studies in Nonlinear Dynamics and Econometrics.

Li, Li, and Shi (2017); Factor models for multivariate time series, see Pan and Yao (2008), Li and Pan (2008), Lam, Yao, and Bathia (2011), Lam and Yao (2012), Xia, Xu, and Zhu (2015), Chan, Lu, and Yau (2017), Xia, Liang, Wu, and Wong (2018) and Xia, Wong, Shen, and He (2022); and So-called dynamic factor models, see Forni, Wallin, Lippi, and Reichlin (2000), Hallin and Liska (2007), Amengual and Watson (2007), Bai and Ng (2007) and Onatski (2009), among others. Some of these references assume that factors are given and subsequently estimate factor loadings. But many papers assume factors are unknown and should be determined based on data. If we try to find the hidden unknown factors, one of the key problems is how to estimate the number of factors. In recent years, several estimation methods have been developed for the number of factors in various factor models, and we are particularly interested in the method proposed in Ahn and Horenstein (2013) for the determination of factor number in a static approximate factor model.

For the static approximate factor models, Bai and Ng (2002) estimated the number of factors by minimizing the information criteria (IC) for model selection. In practice, however, this criteria has been known to overestimate the number of factors. Onatski (2010) considered the so-called "Edge Distribution" estimator by using the differences between eigenvalues of sample covariance matrix. As reported by Ahn and Horenstein (2013), although the methods proposed by Bai and Ng (2002) and Onatski (2010) performed well in a model with independent idiosyncratic errors, they have unsatisfactory finite sample properties in the case with cross-sectional dependent errors. Therefore, two eigenvalue ratio (ER) estimators were proposed and shown to be better when the idiosyncratic errors are cross-sectionally correlated or serially correlated in Ahn and Horenstein (2013). Similarly Wu (2016) used a ratio-type method based on the ratio of transformation function of two adjacent eigenvalues. As a modification to the ER and GR (growth ratio) estimators in Ahn and Horenstein (2013), Xia, Liang, and Wu (2017) proposed an estimator obtained by maximizing the ratio of two adjacent transformed contribution of the eigenvalues, and Wu (2018) obtained an estimator by maximizing the difference between function values of two adjacent eigenvalues.

To determine the number of factors in high-dimensional time series, Lam and Yao (2012) proposed a ratio-type ER estimator based on the ratio of two estimated eigenvalues. Xia et al. (2015) argued that the ratio-type estimators may suffer from the instability of the $0/0$ type ratio values. As a modification of the ER estimator, a ridge-type ratio estimator was proposed by adding a so-called ridge parameter to the numerator and denominator of the ratio of two adjacent eigenvalues in their paper. Furthermore, Xia et al. (2018) suggested a ratio-type estimator by minimizing the ratio of the contributions of two adjacent eigenvalues.

It is well known that a factor model is useful when: the dimension of common factor vector is significantly lower than the dimension of response vector; and

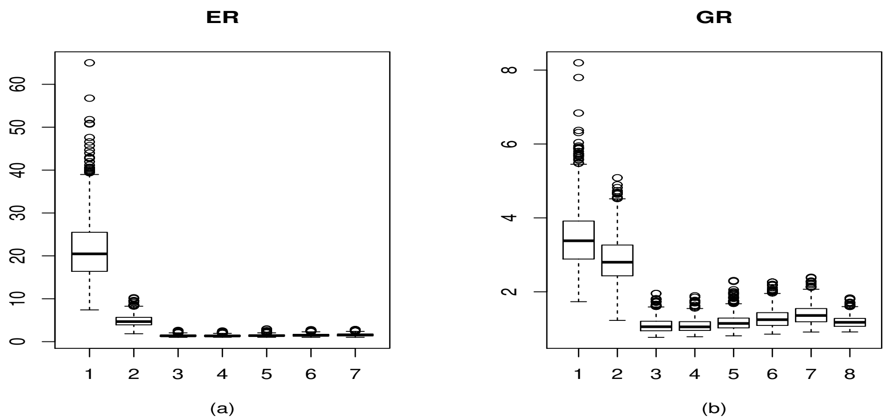

Figure 1: The correct rates of ER and GR in Ahn and Horenstein (2013) for model (3.1) with $\theta = 1$, $kmax = 8$, $\rho = 0.5$, $\beta = 0.2$, $J = \max\{10, N/20\}$ and two factors $F_1 \sim N(0,1)$, $F_2 \sim N(0,20)$, assuming there is a dominant factor and the errors are both serially and cross-sectionally correlated.

at the same time, the cumulative contribution rate of factors is high sufficiently as well. Then, the dimension reduction is achieved because the high-dimensional vector of response variables is driven by a lower-dimensional vector of common factors. The cumulative contribution rate measures the explanatory power of common factors, and thus is important in factor models. However, the eigenvalue ratio-type estimators are determined only according to the mutation of two adjacent eigenvalues, without considering the cumulative contribution rate of all factors which represents the driving power of the common factors in the factor models. Therefore, although the ER and GR estimators of Ahn and Horenstein (2013) are asymptotically consistent when the sample size is large, they often underestimate the true number of factors in finite sample cases, especially when there is a dominant factor and the errors are both serially and cross-sectionally correlated, as the simulation example in Figure 1 shows.

In this paper, we propose a different method based on further sample information for determining the number of factors in static approximate factor models. The proposed estimators are obtained by maximizing the ratio of two adjacent eigenvalues and the cumulative contribution rate of the total common factors. Our estimators are proved to be asymptotically consistent like the ER and GR estimators of Ahn and Horenstein (2013) under the same mild conditions. But Monte Carlo simulation experiments show that the proposed approaches have higher correct rates in the finite-sample cases as shown in Figure 2. That is, by numerical

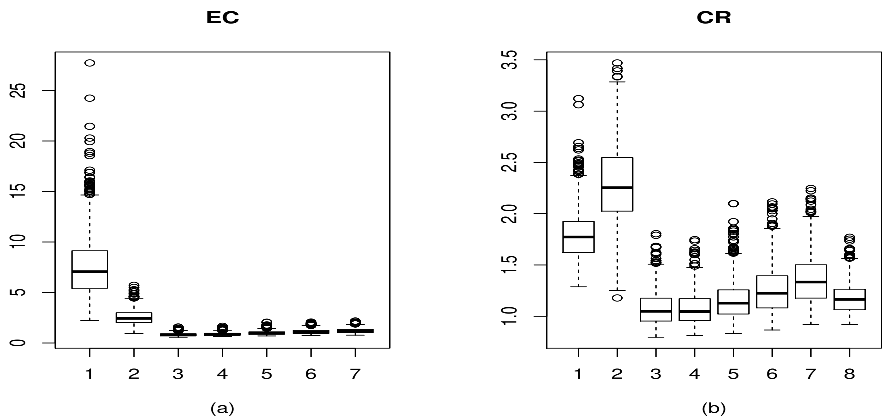

Figure 2: The correct rates of EC and CR proposed in this paper for model (3.1) below with $\theta = 1$, $kmax = 8$, $\rho = 0.5$, $\beta = 0.2$, $J = \max\{10, N/20\}$, and two factors $F_1 \sim N(0,1)$, $F_2 \sim N(0,20)$, assuming there is a dominant factor and the errors are both serially and cross-sectionally correlated.

comparison with the estimators of Ahn and Horenstein (2013), our estimators have superior finite-sample properties.

The rest of this paper is organized as follows: The methodological development and asymptotic properties of the proposed estimators are described in Section 2; Simulation studies are presented in Section 3; Section 4 gives an example on real economic data; The concluding remarks are given in Section 5; All the proofs of theoretical asymptotic properties are given in the Appendix.

2 Methodology and asymptotic properties

Consider the approximate factor model of Chamberlain and Rothschild (2014). Let x_{it} denote the response variable $i(= 1, ..., N)$ at time $t(= 1, ..., T)$. The variables are generated by an $r \times 1$ vector of common factors f_t as follows

$$x_{.t} = \Lambda f_t + \varepsilon_{.t},\ t = 1, ..., T, \tag{1}$$

where $x_{.t} = (x_{1t}, ..., x_{Nt})'$ is an $N \times 1$ vector of response variables, $\Lambda = (\lambda_1, ..., \lambda_N)'$, in which λ_i is an r-dimensional vector of factor loadings for variable i, and $\varepsilon_{.t} = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ is the $N \times 1$ vector of the idiosyncratic components of response variables. The factors, factor loadings, and idiosyncratic components are not observed.

Let $X = (x_{.1}, ..., x_{iT})'$, $F = (f_1, f_2, ..., f_T)'$, and $E = (\varepsilon_{.1}, ..., \varepsilon_{.T})'$. Then, model (2.1) can be rewritten as the matrix form

$$X = F\Lambda' + E. \tag{2}$$

We treat the entries in Λ as parameters and those in F as random variables.

Throughout this paper, A' denotes the transpose of a matrix A. $\|A\| = [trace(A'A)]^{1/2}$ means the norm of a matrix A. $\psi_k(A)$ denotes the kth largest eigenvalues of a positive semidefinite matrix A. Two scalars c_1 and c_2 denote generic positive constants. For any real z, $[z]$ means the integer part of z. The notation $a \asymp b$ denotes $a = O_p(b)$ and $b = O_p(a)$. Finally, we denote $m = \min\{N, T\}$ and $M = \max\{N, T\}$.

Let

$$\tilde{\mu}_{NT,k} = \psi_k[XX'/(NT)] = \psi_k[X'X/(NT)], k = 1, ..., m. \tag{3}$$

denote the kth largest eigenvalues of the matrix $XX'/(NT)$. Ahn and Horenstein (2013) proposed two ratio-type estimators, i.e. ER and GR estimators, to determine the number of common factors as follows:

$$\hat{r}_{ER} = \arg \max_{1\le k\le kmax} \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}}, \ \hat{r}_{GR} = \arg \max_{1\le k\le kmax} \frac{\ln[V(k-1)/V(k)]}{\ln[V(k)/V(k+1)]}, \tag{4}$$

where

$$V(k) = \sum_{j=k+1}^{m} \tilde{\mu}_{NT,j} \tag{5}$$

and *kmax* is the predetermined possible maximum value of the number of factors. Ahn and Horenstein (2013) proved that the ER and GR estimators are consistent when $m \to \infty$.

In order to avoid the possibility of underestimating ER and GR estimators, in this paper, we consider determining the number of factors in model (2.1) by the following forms.

$$\begin{aligned} \tilde{r} &= \arg \min_{1\le k\le kmax} \left\{ \frac{\tilde{\mu}_{NT,k+1}}{\tilde{\mu}_{NT,k}} \frac{V(k)}{V(0)} \right\} = \arg \min_{1\le k\le kmax} \left\{ \frac{\tilde{\mu}_{NT,k+1}}{\tilde{\mu}_{NT,k}} V(k) \right\} \\ &= \arg \max_{1\le k\le kmax} \left\{ \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} \frac{1}{V(k)} \right\}, \end{aligned} \tag{6}$$

where $V(0) = trace(XX')/(NT)$. Note that

$$\frac{V(k)}{V(0)} = 1 - \frac{1}{V(0)} \sum_{j=1}^{k} \tilde{\mu}_{NT,j}.$$

The second term in the right side of the above equation is the cumulative contribution rate of the first k largest eigenvalues of the matrix $XX'/(NT)$. Therefore, if the number of factors in model (2.1) is $r = \tilde{r}$, then the cumulative contribution rate of the total common factors for the response variables is $\sum_{j=1}^{\tilde{r}} \tilde{\mu}_{NT,j}/V(0)$. Thus, the resulting estimator $\tilde{r}$ can give consideration comprehensively to both the ratio of two adjacent eigenvalues and the cumulative contribution rate of factors. However, it is observed that $V(k)$ descends with k. The estimator $\tilde{r}$ obtained from (2.6) may be an overestimate. Moreover, it can be seen that the consistency of the estimator $\tilde{r}$ can not be ensured when $m \to \infty$. To avoid these issues, we suggest a ridge-type estimator based on the forms (2.5), as a modification of the ER estimator, given by

$$\hat{r}_{EC} = \arg \max_{1 \le k \le kmax} \left\{ \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} \frac{1}{v + V(k)} \right\}, \tag{7}$$

where $v > 0$ is a pre-specified positive value. We suggest that it is chosen as $v = O_p(1)$. It is seen from Corollary 1 that if $v = 1$, the asymptotic property of $\hat{r}_{EC}$ is satisfactory. Here, the term EC refers to "Eigenvalue ratio" and "Cumulative contribution rate". This is an idea similar to adding a ridge parameter v to $V(k)$ in the last term of (2.6). Intuitively, the EC estimator $\hat{r}_{EC}$ should work better than the ER estimator $\hat{r}_{ER}$, because the former uses more information from the sample.

Furthermore, we wish to consider the contribution rates sufficiently and propose using the ratio of two adjacent contribution rates to determine the number of factors as follows

$$\hat{r}_{CR} = \arg \max_{1 \le k \le kmax} \left\{ \frac{\tilde{\mu}_{NT,k}/V(k-1)}{\tilde{\mu}_{NT,k+1}/V(k)} \right\}, \tag{8}$$

where $V(k)$ is defined by (2.5). This estimator will prove to be asymptotically consistent as well, see Theorem 2 below. Simulation studies show the estimator defined by (2.8) has better finite sample properties than the GR estimator in Ahn and Horenstein (2013).

Remark 1. It follows from Lamma A.12 of Ahn and Horenstein (2013) that if r is the true value of the number of common factors in model (2.1), then $V(r+1) \asymp O_p(1)$. Thus, in practice, we suggest choosing v in (2.7) as $\hat{v} = dV(\hat{r}_{ER}+1)$, where $d > 0$ is a constant, $V(\hat{r}_{ER}+1) = trace(XX')/(NT) - \sum_{k=1}^{\hat{r}_{ER}} \tilde{\mu}_{NT,k}$, where $\hat{r}_{ER}$ is the ER estimator obtained from (2.4). In fact, it can be seen that the new estimators introduce two different multipliers compared to the ER estimator, i.e., $EC(k) = ER(k)\frac{V(k)}{V(k-1)}$ and $CR(k) = ER(k)\frac{1}{\upsilon+V(k)}$. Note that $\frac{1}{\upsilon+V(k)} = O_p(1)$ and $\frac{V(k)}{V(k-1)} = O_p(1)$ are needed to justify the consistency holding for $\hat{r}_{EC}$ and $\hat{r}_{CR}$. The new estimators are consistent because the additional multipliers do not change their asymptotic behavior compared to ER, i.e., both $\hat{r}_{EC}$ and $\hat{r}_{CR}$ are diverging at the rate

$\min(N,T)$ for $k=r$ and remain bounded for $k\neq r$. Moreover, the main part

$$\tilde{\mu}_{NT,k}/V(k-1)=\tilde{\mu}_{NT,k}/\sum_{i=k}^{m}\tilde{\mu}_{NT,i}$$

in *CR* estimator, which measures $\tilde{\mu}_{NT,k}$'s influence on $V(k-1)$, and can be called the contribution of $\tilde{\mu}_{NT,k}$ for $V(k-1)$. As the modifications to ER and GR estimators, we then name them as "both the ratio of two adjacent eigenvalues and the cumulative contribution rate of factors" and "contribution rate ratio" (hereafter EC and CR) estimators.

For the consistency of the estimator $\hat{r}_{EC}$, we impose some assumptions on the factor model (2.1). These assumptions are the same as those in Ahn and Horenstein (2013).

Assumption A: (i) Let $\mu_{NT,k}=\psi_k[(\Lambda'\Lambda/N)(F'F/T)]$ for $k=1,...,r$. Then. for each $k=1,...,r$, $p\lim_{m\to\infty}\mu_{NT,k}=\mu_k$, and $0<\mu_k<\infty$. (ii) r is finite.

Assumption B: (i) $E\|f_t\|^4<c_1$ and $\|\lambda_i\|<c_2$ for all t and i.

(ii) $E(\|N^{-1/2}\sum_i\varepsilon_{it}\lambda_i\|^2)<c_1$ for all t.

(iii) $E(N^{-1}\sum_i\|T^{-1/2}\sum_t\varepsilon_{it}f_t\|^2)=E[(NT)^{-1}\|E'F\|^2]<c_1$.

Assumption C: (i) $0<y\equiv\lim_{m\to\infty}m/M\leq 1$. (ii) $E=R_T^{1/2}UG_N^{1/2}$, where $U'=(u_{it})_{N\times T}$, and $R_T^{1/2}$ and $G_N^{1/2}$ are the symmetric square roots of $T\times T$ and $N\times N$ positive semidefinite matrices R_T and G_N, respectively. (iii) The u_{it} are independent and identically distributed (i.i.d.) random variables with uniformly bounded moments up to the fourth order. (iv) $\psi_1(R_T)<c_1$ and $\psi_1(G_N)<c_1$, uniformly in T and N, respectively.

Assumption D: (i) $\psi_T(R_T)>c_2$ for all T. (ii) Let $y^*=\lim_{m\to\infty}m/N=\min\{y,1\}$. Then, there exists a real number $d^*\in(1-y^*,1]$ such that $\psi_{[d^*N]}(G_N)>c_2$ for all N.

Remark 2. Assumptions C and D are sufficient, but not necessary, conditions for our results. Weaker conditions sufficient for our results are

$$\psi_1(EE'/M)\asymp O_p(1), \tag{9}$$

$$\psi_{[d^cm]}(EE'/M)]\geq c+o_p(1), \tag{10}$$

for some positive and finite real number c and some $d^c\in(0,1]$. The condition (9) rules out the possibility that the error matrix E contains common factors. The condition (10) indicates that the first largest $[d^cm]$ eigenvalues of EE'/M are bounded away from zero. It follows from Lemma A.9 of Ahn and Horenstein (2013) that the Assumptions C and D are sufficient for both (9) and (10).

Our main results are as follows.

Theorem 1. If Assumptions A-D hold with $r \geq 1$, then there exists $d^c \in (0,1]$ such that $\lim_{m\to\infty} Pr(\hat{r}_{EC} = r) = 1$, for any $kmax \in (r, [d^c m] - r - 1]$.

Theorem 2. Suppose that Assumptions A-D hold with $r \geq 1$. Then, there exists $d^c \in (0,1]$ such that $\lim_{m\to\infty} Pr(\hat{r}_{CR} = r) = 1$, for any $kmax \in (r, [d^c m] - r - 1]$.

The proofs of Theorem 1 and Theorem 2 are presented in the Appendix. It can be seen from the proof that if $v = 1$ the assertion from Theorem 1 holds. Thus, we have the following corollary.

Corollary 1. Let $v = 1$ in the definition of $\hat{r}_{EC}$ given by (2.7), and suppose that Assumptions A-D hold with $r \geq 1$. Then, there exists $d^c \in (0,1]$ such that $\lim_{m\to\infty} Pr(\hat{r}_{EC} = r) = 1$, for any $kmax \in (r, [d^c m] - r - 1]$.

In the case with no factors, i.e. $r = 0$, we can modify slightly the above estimator $\hat{r}_{EC}$ using the same method as Ahn and Horenstein (2013). Specifically, we can defined a mock eigenvalue, e.g. $\tilde{\mu}_{NT,0} = V(0)/\ln(m)$, such that

$$\tilde{\mu}_{NT,0} \to 0, \quad m\tilde{\mu}_{NT,0} \to \infty, \tag{11}$$

as $m \to \infty$, although a finite multiple of the choice is also available. Then, we have the following corollary.

Corollary 2. Suppose that Assumptions A-D hold with $r \geq 0$, and redefine $\hat{r}_{EC}$ using $\tilde{\mu}_{NT,0}$ satisfying (11) for $k = 0$. Then $\lim_{m\to\infty} Pr(\hat{r}_{EC} = r) = 1$.

Corollary 2 can be implied by (11) and Lemma 2 in the appendix as follows: When $r = 0$, from Lemma 2 we have $\tilde{\mu}_{NT,j} = O_p(m^{-1})$, $j = 1,2,...,kmax$, which implies $\frac{\tilde{\mu}_{NT,j}}{\tilde{\mu}_{NT,j+1}} = O_p(1)$, $j = 1,2,...,kmax$, but from (11), $\frac{\tilde{\mu}_{NT,0}}{\tilde{\mu}_{NT,1}} = O_p(1)m\tilde{\mu}_{NT,0} \to \infty$. When $r \geq 1$, $\frac{\tilde{\mu}_{NT,0}}{\tilde{\mu}_{NT,1}} = O_p(1)\tilde{\mu}_{NT,0} \to 0$.

3 Simulation studies

In this section, extensive simulation experiments are carried out to investigate the performance of the proposed estimators EC and CR. To compare with competitors, our simulation exercises will be carried on the model of Ahn and Horenstein (2013):

$$x_{it} = \sum_{j=1}^{r} \lambda_{ij} f_{jt} + \sqrt{\theta} u_{it}, \quad u_{it} = \sqrt{\frac{1-\rho^2}{1+2J\beta^2}} e_{it}, \tag{12}$$

where $e_{it} = \rho e_{i,t-1} + v_{it} + \sum_{l=\max\{i-J,1\}}^{i-1} \beta v_{lt} + \sum_{l=i+1}^{\min\{i+J,n\}} \beta v_{lt}$, and λ_{ij} and v_{lt} are both drawn from the standard normal distribution $N(0,1)$ independently. The factors f_{jt} are drawn from normal distributions with zero means. Note that the idiosyncratic components (errors) u_{it} have been normalized to have variance 1.

In our simulation experiments, the settings of parameters are important to show the various structures of data. Control parameter θ is the inverse of the signal-to-noise ratio (*SNR*) of each factor, i.e., $SNR = 1/\theta = \mathrm{var}(f_{jt})/\mathrm{var}(\sqrt{\theta} u_{it})$ with $\mathrm{var}(f_{jt}) = 1$. When it is necessary to change *SNR*s of all factors, we adjust the value of θ while fixing variances of factors at 1. To change the *SNR* of a single factor, we adjust the variance of the factor with θ fixed at 1. Moreover, ρ can show the serial correlation of the idiosyncratic errors over the time index, and β and J can control the cross-sectional dependency of the idiosyncratic errors over the individual index.

Our simulations are classified into four parts with the same settings as Ahn and Horenstein (2013). Performances of our *EC and CR* estimators are compared with their counter parts: ER and GR estimators in Ahn and Horenstein (2013), BIC_3 estimator in Bai and Ng (2002), ED estimator in Onatski (2010), and IC_1 estimator in Alessi et al. (2010). The estimation procedure for the number r is replicated 1000 times, and the corresponding results are denoted by $x(y|z)$, where x is the number of different estimators in the replications, the numbers y and z in parentheses $(y|z)$ are the number of under- and over-estimations in the 1000 replications, respectively.

The first part is designed to investigate four examples with different error covariance structures which will influence the finite-sample performances of *EC* and *CR* estimators as well as other estimators:

Example 1. Independent and identically distributed errors $(\rho = \beta = J = 0)$;
Example 2. Serially correlated errors $(\rho = 0.7, \beta = J = 0)$;
Example 3. Cross-sectionally correlated errors $(\rho = 0, \beta = 0.5, J = \max\{10, \frac{N}{20}\})$;
Example 4. Both serially and cross-sectionally correlated errors $(\rho = 0.5, \beta = 0.2, J = \max\{10, \frac{N}{20}\})$.

Table 1 reports the simulation results on the effects of the model structure on performance of seven estimators when three factors are drawn from $N(0,1)$ and θ is fixed at 1. In example 1 with i.i.d. errors, BIC_3, IC_1 and *CR* are almost better than the other estimators with $N \leq 50$; when $N \geq 50$, except *ED* estimator, all estimators have the same desired performance. In example 2 with serially correlated errors, *CR* and *GR* are slightly better than other estimators with $N, T \leq 50$; when $N \geq 50$ with growing T, *EC*, *ER* and BIC_3 estimators improve as well as *CR* and *GR* estimators, while $T \leq 50$, BIC_3 estimator gives poor performance. In example 3 with cross-sectionally correlated errors, all estimators have bad performance when $N \leq 50$; when $N = 100, T \leq 100$, *ER*, *GR* and *CR* outperform other estimators; when $N = 100, T = 200$, *ED* estimator has the best performance; when $N = 200, T \leq 25$, *CR* estimator has the best performance; when $N = 200, T = 25$, *CR* estimator has the best performance; when $N = 200, T \geq 50$, *ER*, *GR*, *EC* and *CR* almost show the same satisfactory performance. Meanwhile, BIC_3 always performs poorly. In example 4 with serially/ cross-sectionally correlated errors, ER is superior to other estimators with $N = 25$, the second best one is EC estimator; when $N = 50, T \leq 100$, EC performs better than other estimators; when $N \geq 100, T = 25$, CR estimator is the best one; when $N = 2000, T \geq 100$, BIC_3 estimator becomes as good as ER, GR, EC and CR estimators. It appears that the performance of BIC_3 estimator is much more sensitive to cross-sectional correlation than to autocorrelation in the errors from examples 3 and 4.

Table 1: Performances of seven estimators for model (3.1) with three factors and I.I.D. errors with $\theta = 1, kmax = 8, \rho = \beta = J = 0$

N	T	r	ER	GR	BIC_3	IC_1	ED	EC	CR
25	25	3	888(112|0)	948(52|0)	838(0|162)	986(13|1)	982(3|15)	924(76|0)	964(36|0)
	50	3	984(16|0)	995(5|0)	1000(0|0)	1000(0|3)	993(0|7)	994(6|0)	997(3|0)
	100	3	996(4|0)	998(2|0)	1000(0|0)	1000(0|0)	993(0|7)	996(4|0)	998(2|0)
	200	3	999(1|0)	1000(0|0)	999(1|0)	1000(0|0)	997(0|3)	999(1|0)	1000(0|0)
50	25	3	988(12|0)	997(3|0)	1000(0|0)	998(0|2)	987(0|13)	989(11|0)	997(3|0)
	50	3	1000(0|0)	1000(0|0)	1000(0|0)	999(0|1)	990(0|10)	1000(0|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	996(0|4)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	992(0|8)	1000(0|0)	1000(0|0)
100	25	3	994(6|0)	999(1|0)	1000(0|0)	1000(0|0)	993(0|7)	997(3|0)	1000(0|0)
	50	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	999(0|1)	1000(0|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	997(0|3)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	992(0|8)	1000(0|0)	1000(0|0)
200	25	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	993(0|7)	1000(0|0)	1000(0|0)
	50	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	991(0|9)	1000(0|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	994(0|6)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)	996(0|4)	1000(0|0)	1000(0|0)

Table 2: Performances of seven estimators for model (3.1) with three factors and serially correlated errors with $\theta = 1, kmax = 8, \rho = 0.7, \beta = J = 0$

N	T	r	ER	GR	BIC_3	IC_1	ED	EC	CR
25	25	3	612(314|74)	629(200|171)	0(0|1000)	521(242|237)	570(132|298)	632(137|231)	624(170|206)
	50	3	878(122|0)	927(66|7)	5(0|995)	869(35|96)	894(16|90)	929(64|7)	938(51|11)
	100	3	974(26|0)	991(9|0)	963(0|37)	965(3|32)	971(0|29)	982(18|0)	994(6|0)
	200	3	994(6|0)	998(2|0)	1000(0|0)	998(0|2)	981(0|19)	997(3|0)	998(2|0)
50	25	3	785(198|17)	812(109|79)	0(0|1000)	620(82|298)	719(72|209)	806(85|109)	810(90|100)
	50	3	981(19|0)	993(6|1)	31(0|969)	885(0|115)	942(0|58)	991(19|0)	993(6|1)
	100	3	1000(0|0)	1000(0|0)	998(0|2)	963(0|37)	971(0|29)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	998(0|2)	990(0|10)	1000(0|0)	1000(0|0)
100	25	3	864(130|6)	907(61|32)	0(0|1000)	623(15|362)	823(51|126)	893(55|52)	905(50|45)
	50	3	997(3|0)	998(2|0)	68(0|932)	953(0|47)	989(0|11)	999(1|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	1000(0|0)	991(0|9)	990(0|10)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	996(0|4)	989(0|11)	1000(0|0)	1000(0|0)
200	25	3	909(91|0)	945(46|9)	0(0|1000)	557(8|435)	920(24|56)	943(46|11)	955(31|14)
	50	3	1000(0|0)	1000(0|0)	86(0|914)	966(0|34)	997(0|3)	1000(0|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	1000(0|0)	995(0|5)	995(0|5)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	999(0|1)	994(0|6)	1000(0|0)	1000(0|0)

Table 3: Performances of seven estimators for model (3.1) with three factors and cross-sectionally correlated errors with
$\theta = 1, kmax = 8, \rho = 0, \beta = 0.5, J = \max\{10, N/20\}$

N	T	r	ER	GR	BIC_3	IC_1	ED	EC	CR
25	25	3	4(2|994)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	2(0|998)	0(0|1000)
	50	3	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)
	100	3	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)
	200	3	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)
50	25	3	69(31|900)	12(6|982)	0(0|1000)	0(0|1000)	0(0|1000)	22(1|977)	10(3|987)
	50	3	31(2|967)	3(0|997)	0(0|1000)	0(0|1000)	0(0|1000)	4(0|996)	2(0|998)
	100	3	16(0|984)	1(0|999)	0(0|1000)	0(0|1000)	0(0|1000)	1(0|999)	0(0|1000)
	200	3	4(0|996)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)	0(0|1000)
100	25	3	755(160|0)	737(73|190)	0(0|1000)	645(116|239)	696(111|193)	702(36|262)	722(66|212)
	50	3	947(46|7)	945(18|37)	0(0|1000)	862(32|106)	922(25|53)	931(15|54)	943(15|42)
	100	3	998(2|0)	999(1|0)	0(0|1000)	982(0|18)	994(0|6)	994(0|6)	999(1|0)
	200	3	999(1|0)	999(1|0)	0(0|1000)	858(0|142)	1000(0|0)	999(1|0)	999(1|0)
200	25	3	941(58|1)	969(30|1)	0(0|1000)	850(2|48)	934(5|61)	963(35|2)	975(23|2)
	50	3	998(2|0)	1000(0|0)	0(0|1000)	924(0|76)	984(0|16)	1000(0|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	0(0|1000)	985(0|15)	991(0|9)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	0(0|1000)	1000(0|0)	1000(0|0)	1000(0|0)	1000(0|0)

Table 4: Performances of seven estimators for model (3.1) with three factors and serially/cross-sectionally correlated errors with
$\theta = 1, kmax = 8, \rho = 0.5, \beta = 0.2, J = \max\{10, N/20\}$

N	T	r	ER	GR	BIC_3	IC_1	ED	EC	CR
25	25	3	543(271|186)	510(150|340)	0(0|1000)	388(95|527)	385(76|559)	544(121|335)	492(119|389)
	50	3	640(160|200)	576(51|373)	2(0|998)	207(14|779)	222(11|767)	632(49|319)	544(37|419)
	100	3	722(100|178)	623(18|359)	10(0|990)	46(2|952)	64(0|936)	687(28|285)	566(13|421)
	200	3	772(79|149)	599(17|384)	20(0|980)	0(0|1000)	3(0|997)	725(29|246)	510(12|478)
50	25	3	790(164|46)	808(78|114)	0(0|1000)	532(25|443)	624(41|345)	816(76|108)	802(65|133)
	50	3	958(29|13)	956(6|38)	28(0|972)	303(0|697)	580(1|419)	964(9|27)	945(4|51)
	100	3	996(3|1)	998(0|2)	31(0|969)	95(0|905)	436(0|564)	997(1|2)	993(0|7)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	999(0|1)	986(0|14)	1000(0|0)	1000(0|0)
100	25	3	931(67|2)	960(34|6)	59(0|941)	746(4|50)	863(9|28)	956(40|4)	967(26|7)
	50	3	1000(0|0)	1000(0|0)	455(0|545)	731(0|269)	893(0|107)	1000(0|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	568(0|452)	654(0|346)	913(0|87)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	658(0|342)	751(0|249)	967(0|33)	1000(0|0)	1000(0|0)
200	25	3	968(32|0)	990(10|0)	548(0|452)	857(0|143)	952(0|48)	984(16|0)	993(6|1)
	50	3	1000(0|0)	1000(0|0)	994(0|6)	872(0|128)	954(0|46)	1000(0|0)	1000(0|0)
	100	3	1000(0|0)	1000(0|0)	1000(0|0)	902(0|98)	971(0|29)	1000(0|0)	1000(0|0)
	200	3	1000(0|0)	1000(0|0)	1000(0|0)	948(0|52)	991(0|9)	1000(0|0)	1000(0|0)

The second part examines the effects of weak factors on the estimators. We consider two cases: the case in which all three factors have weak explanatory power ($SNR = 0.17$), and the case in which two factors are strong ($SNR = 1$) and one factor is weak ($SNR < 1$). Table 5 reports the results for the first case from our simulations. In short, when $N, T \leq 125$, none of the estimators are satisfactory; when $N, T = 150$, the ER estimator is superior to other estimators; when $N, T \geq 175$, the CR estimator outperforms other estimators. Table 6 reports the results for the second case from our experiments. When $SNR_3 \leq 0.25$, the IC_1 estimator shows better than others, but all estimators seem to have poor performance; when $0.3 \leq SNR_3 \leq 0.45$, the EC performs best; when $SNR_3 = 0.5$, the CR estimator becomes the most suitable one.

Table 5: Effects of all weak factors with serially/cross-sectionally correlated errors with $kmax = 8, \theta = 6, \rho = 0.5, \beta = 0.2, J = \max\{10, N/20\}$, and $F_1, F_2, F_3 \sim N(0,1)$

N	T	r	ER	GR	BIC_3	IC_1	ED	EC	CR
25	25	3	139(503|358)	130(379|491)	0(0|1000)	102(602|296)	117(466|417)	32(12|956)	134(354|512)
50	50	3	44(163|793)	29(82|889)	0(0|1000)	16(217|767)	7(479|514)	1(0|999)	25(68|907)
75	75	3	57(516|12)	11(28|961)	0(0|1000)	8(497|495)	2(617|381)	0(0|1000)	11(25|964)
100	100	3	100(285|615)	45(147|808)	0(0|1000)	7(849|144)	7(974|19)	0(0|1000)	44(142|814)
125	125	3	211(431|358)	156(313|531)	0(0|1000)	44(885|71)	33(964|3)	1(0|999)	153(303|544)
150	150	3	419(381|300)	396(305|299)	0(0|1000)	190(752|58)	104(889|7)	39(2|959)	394(299|307)
175	175	3	668(262|70)	683(211|106)	0(0|1000)	523(91|435)	380(614|6)	235(1|764)	685(206|109)
200	200	3	877(106|17)	902(71|27)	0(0|1000)	840(137|23)	731(267|2)	613(3|384)	903(70|27)

Table 6: Effects of strong and weak factors with serially/cross-sectionally correlated errors with $N = T = 100, \theta = 1, kmax = 8, \rho = 0.5, \beta = 0.2, J = \max\{10, N/20\}$, $F_1, F_2 \sim N(0,1)$, and $F_3 \sim N(0, SNR_3)$.

SNR_3	r	ER	GR	BIC_3	IC_1	ED	EC	CR
0.10	3	0(1000|0)	1(999|0)	0(0|1000)	72(743|185)	28(943|29)	1(998|1)	1(998|1)
0.15	3	1(999|0)	1(999|0)	0(0|1000)	73(676|251)	36(939|25)	0(998|2)	2(998|0)
0.20	3	5(995|295)	8(988|4)	0(0|1000)	188(505|207)	131(827|42)	7(984|9)	10(985|5)
0.25	3	37(963|0)	93(898|7)	0(0|1000)	393(269|338)	344(614|42)	83(878|39)	106(875|19)
0.30	3	163(836|1)	265(712|23)	0(0|1000)	489(150|361)	573(378|195)	254(694|52)	300(665|35)
0.35	3	377(623|0)	533(445|22)	0(0|1000)	604(35|361)	779(172|49)	501(419|80)	570(400|30)
0.40	3	582(417|1)	710(254|36)	0(0|1000)	589(14|397)	860(97|43)	685(234|81)	747(203|50)
0.45	3	745(254|1)	862(117|21)	0(0|1000)	602(7|391)	926(39|35)	832(115|53)	874(99|27)
0.50	3	877(123|0)	935(55|10)	0(0|1000)	593(4|403)	944(19|37)	918(50|32)	945(43|12)

The third part of our simulations considers the case in which two factors have strong explanatory power, but one of factor powers is increasingly dominant. We generate data using two factors with different SNRs ($SNR_1 = 1, SNR_2 > 1$), i.e., the two factors are drawn from $N(0,1)$ and $N(0, SNR_2)$, respectively. We investigate four cases $N = T = 75$; $N = T = 100$; $N = T = 150$; $N = T = 200$.

Table 7 reports the results of six estimators when there is one dominant factor. For such a case with large SNR_2, when $N \leq 150$, the CR estimator are outperforming other estimators in almost all cases.

Table 7: Effects of dominant factor with two factors with $\theta = 1$, $kmax = 8$, $\rho = 0.5$, $\beta = 0.2$, $J = \max\{10, N/20\}$, $F_1 \sim N(0,1)$, and $F_2 \sim N(0, SNR_2)$

$N=T$	SNR_2	r	ER	GR	BIC_3	IC_1	ED	EC	CR
75	1	2	987(13|0)	961(1|38)	0(0|1000)	36(0|964)	394(0|606)	909(1|90)	931(1|68)
	3	2	788(212|0)	950(41|9)	0(0|1000)	28(0|972)	358(1|641)	919(1|90)	962(13|25)
	7	2	210(790|0)	757(243|0)	0(0|1000)	36(0|964)	357(0|643)	503(496|1)	887(97|16)
	10	2	76(924|0)	617(383|0)	0(0|1000)	34(0|966)	380(0|620)	281(719|90)	902(76|22)
	15	2	11(989|0)	477(523|0)	0(0|1000)	28(0|972)	401(0|599)	101(899|0)	844(149|7)
	20	2	1(999|0)	331(669|0)	0(0|1000)	25(0|975)	362(1|637)	42(958|0)	803(192|5)
100	1	2	1000(0|0)	999(0|1)	0(0|1000)	402(0|598)	842(0|158)	1000(0|0)	999(0|1)
	3	2	927(73|0)	994(6|0)	0(0|1000)	401(0|599)	834(0|164)	979(21|0)	999(1|0)
	7	2	372(628|0)	936(64|0)	0(0|1000)	381(0|619)	838(0|162)	704(296|0)	994(6|0)
	10	2	116(884|0)	834(166|0)	0(0|1000)	419(0|581)	851(0|149)	407(593|0)	970(30|0)
	15	2	19(981|0)	693(307|0)	0(0|1000)	376(0|624)	848(0|152)	174(826|0)	968(32|0)
	20	2	4(996|0)	569(431|0)	0(0|1000)	374(0|626)	853(0|147)	61(939|0)	970(30|0)
150	1	2	1000(0|0)	1000(0|0)	0(0|1000)	885(0|115)	989(0|11)	1000(0|0)	1000(0|0)
	3	2	996(4|0)	1000(0|0)	0(0|1000)	891(0|109)	983(0|17)	1000(0|0)	1000(0|0)
	7	2	671(329|0)	998(2|0)	0(0|1000)	872(0|128)	985(0|15)	907(93|0)	1000(0|0)
	10	2	296(704|0)	991(9|0)	0(0|1000)	863(0|137)	984(0|16)	697(303|0)	1000(0|0)
	15	2	58(942|0)	968(32|0)	0(0|1000)	882(0|118)	989(0|11)	355(645|0)	1000(0|0)
	20	2	9(991|0)	943(57|0)	0(0|1000)	861(0|139)	990(0|10)	142(858|0)	999(1|0)
200	1	2	1000(0|0)	1000(0|0)	0(0|1000)	959(0|41)	989(0|11)	1000(0|0)	1000(0|0)
	3	2	1000(0|0)	1000(0|0)	0(0|1000)	964(0|36)	997(0|3)	1000(0|0)	1000(0|0)
	7	2	907(93|0)	1000(0|0)	0(0|1000)	962(0|38)	990(0|10)	988(12|0)	1000(0|0)
	10	2	542(458|0)	1000(0|0)	0(0|1000)	969(31|0)	992(0|8)	890(110|0)	1000(0|0)
	15	2	134(866|0)	999(1|0)	0(0|1000)	955(0|45)	990(0|10)	558(442|0)	1000(0|0)
	20	2	25(975|0)	996(4|0)	0(0|1000)	956(0|44)	989(0|11)	266(734|0)	1000(0|0)

The last part of simulation study investigates how the choice of $kmax$, which can be considered as "smoothing parameters" for the estimators, may influence the performances of estimators. We discuss the effect of large $kmax$ when the eigenvalues $\tilde{\mu}_{NT,i}$ are close to zero for some large $i(< m)$. Six different values are used for $kmax$, and the data generating process is the same as *example* 4 of the first part with $N = T = 150$, the estimated results are recorded in Table 8. The results show that the EC and CR estimators performance nearly the same as the ER and GR estimators, which are insensitive to $kmax$. The ED estimator is less sensitive to $kmax$ than the IC_1 and BIC_3 estimators.

Table 8: Estimation with different values of $kmax$ with $N = T = 150$, $\theta = 1$, $\rho = 0.5$, $\beta = 0.2$, $J = \max\{10, N/20\}$, and $F_1, F_2, F_3 \sim N(0,1)$

kmax	*r*	*ER*	*GR*	BIC_3	IC_1	*ED*	*EC*	*CR*
8	3	1000(0|0)	1000(0|0)	0(0|1000)	889(0|111)	988(0|12)	1000(0|0)	1000(0|0)
12	3	1000(0|0)	1000(0|0)	0(0|1000)	576(0|424)	970(0|30)	1000(0|0)	1000(0|0)
16	3	1000(0|0)	1000(0|0)	0(0|1000)	218(0|782)	896(0|104)	1000(0|0)	1000(0|0)
20	3	1000(0|0)	1000(0|0)	0(0|1000)	235(0|765)	889(0|111)	1000(0|0)	1000(0|0)
25	3	1000(0|0)	1000(0|0)	0(0|1000)	227(0|773)	899(0|101)	1000(0|0)	1000(0|0)
30	3	1000(0|0)	1000(0|0)	0(0|1000)	223(0|777)	874(0|126)	1000(0|0)	1000(0|0)

4 A real data example

An example on the macroeconomic data used by Bernanke, Boivin, and Eliasz (2005) is given as illustration of our methodology. This data set is retrieved from Jean Boivin's webpage and contains 120 monthly macroeconomic variables with 511 time series observations during the period from February 1959 to August 2001. In order to transform the data into stationary time series, we use double-demeaned variables suggested by Ahn and Horenstein (2013) to remove time and time and variable-specific effects from this data. For more detailed information about this data, one can refer to appendix 1 of Bernanke et al. (2005).

The numeric results of the estimated numbers of factors are shown by different estimators as follows. For $kmax = 8, 15, 20$, *ER*, *GR*, *EC*, and *CR* estimates are always 5. But BIC_3, IC_1 and *ED* are affected by *kmax*, because it is found that 8 factors for BIC_3 and IC_1 estimators, and 7 for *ED* estimator when $kmax = 8$; 15 factors for BIC_3 and IC_1 estimators, and 7 for *ED* estimator when $kmax = 15$; 20 factors for BIC_3 and IC_1 estimators, and 16 for *ED* estimator when $kmax = 20$. If we use raw data, we can find that *ER*, *GR*, *EC* and *CR* estimates are always 1, 2, 6 and 7, respectively when $kmax = 8, 15, 20$. Furthermore, BIC_3, IC_1 and *ED* estimators are also found to be sensitive to the choice of *kmax*: BIC_3 and IC_1 estimators are 8, *ED* estimator is 7 with $kmax = 8$; BIC_3 and IC_1 estimators are 15, *ED* estimator is 7 with $kmax = 15$; BIC_1, IC_1 and *ED* estimators are 20 with $kmax = 20$. Thus, in this case without demeaned data, we think that 7 factors may be more reliable.

5 Concluding remarks

In this paper, we introduced two new estimators *EC* and *CR* for the number of factors in static approximate factor models. These new estimators were proved to be convergent to the true number of factors theoretically. Simulation studies showed *EC* and *CR* generally outperform those estimators in existing literature in finite sample cases: They have overcome the potential underestimation of *ER* and *GR*; They are generally insensitive to cross-sectional correlation and autocorrelation;

When all the factors have weak explanatory power, *CR* outperforms other estimators for large N and T, and when factors become strong, *EC* becomes better and *CR* becomes most suitable; *EC* and *CR* are insensitive to choices of *kmax* as *ER* and *GR*. When there is a dominant factor, *CR* always outperforms other estimators.

Acknowledgments

We thank the Editor and the Referee(s) for their insightful comments and suggestions that help us significantly to improve our manuscript. This research was supported by the National Natural Science Foundation of China (No.12171161, 91746102), the Natural Science Foundation of Guangdong Province of China (No. 2022A1515011754), and Ministry of Education in China Project of Humanities and Social Sciences (No.17YJA910002), the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education.

References

Ahn, S. and A. Horenstein (2013): “Eigenvalue ratio test for the number of factors,” *Econometrica*, 81, 1203–1227.

Alessi, L., M. Barigozzi, and M. Capasso (2010): “Improved penalization of determining the number of factors in approximate factor models,” *Statist. Probab. Lett.*, 80, 1806–1813.

Amengual, D. and M. Watson (2007): “Consistent estimation of the number of dynamic factors in a large n and t panel,” *J.Bus. Econom. Statist.*, 25, 91–96.

Bai, J. and S. Ng (2002): “Determining the number of factors in approximate factor models,” *Econometrica*, 70, 191–221.

Bai, J. and S. Ng (2007): “Determining the number of primitive shocks in factor models,” *J. Bus. Econom. Statist.*, 25, 52–60.

Bernanke, B., J. Boivin, and P. Eliasz (2005): “Measuring the effects of monetary policy: A factor augmented vector autoregressive (favar) approach,” *The Quarterly Journal of Economics*, 120, 387–422.

Caner, M. and X. Han (2014): “Selecting the correct number of factors in approximate factor models: the large panel case with group bridge estimators,” *J. Bus. Econom. Statist.*, 32, 359–374.

Chamberlain, G. and M. Rothschild (2014): “Arbitrary, factor structure, and mean-variance analysis on large asset markets,” *Econometrica*, 51, 1281–1304.

Chan, N., Y. Lu, and C. Yau (2017): "Factor modelling for high-dimensional time series: Inference and model selection," *Journal of Time Series Analysis*, 38, 285–307.

Connor, G. and R. Korajzcyk (1993): "A test for the number of factors in an approximate factor model," *J. Finance*, 48, 1263–1291.

Fan, J., Y. Liao, and M. Mincheva (2013): "Large covariance estimation by thresholding principal orthogonal complements," *J. Roy. Statist. Soc. Ser. B*, 75, 603–680.

Forni, M., M. Wallin, M. Lippi, and L. Reichlin (2000): "The generalized dynamic-factor model: idendification and estimation," *Rev. Econ. Stat.*, 82, 540–554.

Hallin, M. and R. Liska (2007): "Determining the number of factors in general dynamic factor model," *J. Amer. Statist. Assoc.*, 102, 603–617.

Lam, C. and Q. Yao (2012): "Factor modeling for high-dimensional time series: Inference for the number of factors," *The Annals of Statistics*, 40, 694–726.

Lam, C., Q. Yao, and N. Bathia (2011): "Estimation for latent factor models for high-dimensional time series," *Biometrika*, 98, 901–918.

Li, H., Q. Li, and Y. Shi (2017): "Determining the number of factors when the number of factors can increase with sample size," *J. Econometrics*, 197, 76–86.

Li, Q. and J. Pan (2008): "Determining the number of factors in a multivariate error correction-volatility factor model," *J. Econometrics*, 12, 45–61.

Onatski, A. (2009): "Testing hypotheses about the number of factors in large factor models," *Ecomometrica*, 77, 1447–1479.

Onatski, A. (2010): "Determining the number of factors from empirical distribution of eigenvalues," *Rev. Econ. Stat.*, 92, 1004–1016.

Pan, J. and Q. Yao (2008): "Modelling multiple time series via common factors," *Biometrika*, 95, 356–379.

Stock, J. and M. Watson (2002): "Forecasting using principal components from a large number of predictors," *J. Amer. Statist. Assoc.*, 97, 1167–1179.

Wu, J. (2016): "Robust determination for the number of common factors in the approximate factor models," *Econom. Lett.*, 144, 102–106.

Wu, J. (2018): "Eigenvalue difference test for the number of common factors in the approximate factor models," *Econom. Lett.*, 169, 63–67.

Xia, Q., R. Liang, and J. Wu (2017): "Transformed contribution ratio test for the number of factors in static approximate factor models," *Computational Statistics and Data Analysis*, 112, 235–241.

Xia, Q., R. Liang, J. Wu, and H. Wong (2018): "Determining the number of factors for high-dimensional time teries," *Statistics and Its Interface*, 11, 307–316.

Xia, Q., H. Wong, S. Shen, and K. He (2022): "Factor analysis for high dimensional time series: consistent estimation and efficient computation," *Statistics Analysis and Data Mining*, 15, 247–263.

Xia, Q., W. Xu, and L. Zhu (2015): "Consistently determining the number of factors in multivariate volatility modelling," *Statistica Sinica*, 25, 1025–1044.