Biometrika (2020), **107**, 2, *pp.* 505–511
Printed in Great Britain

doi: 10.1093/biomet/asz073
Advance Access publication 3 February 2020

A random-perturbation-based rank estimator of the number of factors

BY XINBING KONG

School of Statistics and Mathematics, Nanjing Audit University, 86 West Yushan Road, Nanjing 211815, China
xinbingkong@126.com

SUMMARY

We introduce a random-perturbation-based rank estimator of the number of factors of a large-dimensional approximate factor model. An expansion of the rank estimator demonstrates that the random perturbation reduces the biases due to the persistence of the factor series and the dependence between the factor and error series. A central limit theorem for the rank estimator with convergence rate higher than root n gives a new hypothesis-testing procedure for both one-sided and two-sided alternatives. Simulation studies verify the performance of the test.

Some key words: Large-dimensional approximate factor model; Matrix perturbation; Principal component analysis.

1. INTRODUCTION

Consider the large-dimensional approximate factor model proposed in Chamberlain & Rothschild (1983):

$$(X_t)_{p\times 1} = \Lambda_{p\times r}(f_t)_{r\times 1} + (\epsilon_t)_{p\times 1}, \tag{1}$$

where $X_t = (X_{1t},\ldots,X_{pt})^{\mathrm{T}}$ is a p-dimensional random vector, $\Lambda = (\lambda_{it})_{p\times r}$ is a $p \times r$ factor loading matrix, $f_t = (f_{1t},\ldots,f_{rt})^{\mathrm{T}}$ is an r-dimensional vector of factors with common covariance matrix Σ_f, and $\epsilon_t = (\epsilon_{1t},\ldots,\epsilon_{pt})^{\mathrm{T}}$ is a $p \times 1$ error vector that is cross-sectionally weakly dependent. In recent years a large amount of work has been devoted to estimating the number of factors. Most of these studies are based on separating the diverging and finite eigenvalues in the case where the sample size and dimension tend to infinity simultaneously. To mention a few, Bai & Ng (2002) proposed to estimate the number of factors by minimizing two model selection criterion functions; Onatski (2010) developed an edge distribution estimator by thresholding eigenvalue differences; Ahn & Horenstein (2013) estimated the number of factors by cutting eigenvalue ratios; and Trapani (2018) sequentially tested the divergence of eigenvalues and found a consistent estimate for the number of factors. Variations using high-frequency data include Kong (2017).

In this note, we propose a random-perturbation-based rank estimator of the number of factors. We transform the problem of estimating r into one of estimating the rank of a d-dimensional series of principal components, where d is a preset number larger than r. The role of d is similar to that of $r_{\max}$ in the aforementioned papers. Since the main objective of the present paper is to test or construct a confidence interval for the true number of factors, in practical terms we can choose $d = r_0 + r^*$ for a number r_0 under the null hypothesis, or $d = \hat{r} + r^*$ for some initial estimate $\hat{r}$ such as the Ahn & Horenstein (2013) estimate, with r^* being some positive integer. If $d \leqslant r$, the construction of the rank estimator implies that it is close to d. To avoid underestimating r when a small d is chosen, one could set a series of finite d values and observe when the rank estimator approaches a plateau as d increases.

Using the novel expansion of determinants given in Jacod & Podolskij (2013), we obtain an expansion of the random-perturbation-based rank estimator of the number of factors. This expansion not only shows the consistency of our estimator, but also can be used to assess the accuracy of the estimator in finite samples. Direct use of the expansion yields a central limit theorem for the rank estimator, which is useful for testing hypotheses and constructing confidence intervals for the number of factors. Interestingly, we prove that the rate of convergence of the central limit theorem reaches $p^{-1/4}n^{-1/2}$, which is higher than root n. A simple simulation presented in the Supplementary Material demonstrates that the proposed estimator has a smaller standard error than existing $\sqrt{n}$-consistent estimators.

Compared with existing integer-valued estimates, a distinctive feature of our estimate is that it is real-valued, as it has the form of a ratio of squared determinants. It is precisely this property that makes the central limit theorem possible. The test of Onatski (2009) works well for a right-sided alternative, $r > r_0$ for some number r_0, but it is not applicable to left-sided or two-sided alternatives. On the other hand, to cater to the limiting spectral distribution in random matrix theory, Onatski (2009) assumed $p/n \to c'$ for some constant c'. Based directly on the central limit theorem for the rank estimator, our test is applicable to both one-sided and two-sided alternatives, and requires only that $n = o(p^{3/2})$ as $n, p \to \infty$ simultaneously. As with existing estimators, we only need $n, p \to \infty$ to obtain the consistency of the rank estimator. Connor & Korajczyk (1993) proposed a test assuming observable factors, while here we assume that the factors are latent. Two interesting right-sided rank tests are presented in Cragg & Donald (1997) and Kleibergen & Paap (2006). While these two papers considered testing for the rank of a matrix of fixed dimension, the present paper focuses on a high-dimensional approximate factor model. Another difference is that our rank statistic is based on local expansion of the determinant, while Cragg & Donald (1997) and Kleibergen & Paap (2006) proposed rank statistics of a quadratic form of properly estimated precision matrices.

Related to this paper, Jacod & Podolskij (2013) estimated the maximal rank of the volatility process using a novel perturbation technique. However, their setting differs from ours in several respects. First, Jacod & Podolskij (2013) aimed to estimate the maximal rank of a volatility matrix of fixed dimension, whereas here the target is the number of factors of a large-dimensional approximate factor model. Before the matrix perturbation can be used, we need to reduce the dimension of the factor model and focus on a series of low-dimensional principal components. Second, our model has a large cross-section of idiosyncratic error components, but the multivariate diffusion model in Jacod & Podolskij (2013) can be regarded as a perfect continuous-time factor model without such components. In the present paper, we resort to sample splitting and an orthogonal transform to reduce the idiosyncratic errors. These reduced idiosyncratic errors generate an intrinsic bias of order $\{\min(n, p)\}^{-1/2}$ in the expansion of the squared determinants. This bias matters more than the asymptotic standard error, see the Supplementary Material for details, but we prove that it vanishes in the expansion of the rank estimator. No such technique was introduced in Jacod & Podolskij (2013). Finally, the asymptotic regime of our paper is $n, p \to \infty$ as in a typical high-dimensional time series, while that of Jacod & Podolskij (2013) is of an infill type assuming a fixed time horizon and shrinking time lags. Model (1) above is a static model. The seminal paper by Forni et al. (2000) proposed a generalized dynamic factor model that can accommodate a factor space of infinite dimension, and the factors are loaded via linear filters. Adapting to the dynamic feature, Hallin & Liška (2007) developed an information criterion to estimate the number of factors. In this paper we consider only rank estimation under the static model, and extensions to the generalized dynamic factor model will be left to future work.

Throughout the paper, $\|x\|$ denotes the Euclidean norm of a vector x, $\|X\| = \lambda_{\max}^{1/2}(X^{\mathrm{T}}X)$ is the spectral norm of a matrix X, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue, and $\|X\|_{\mathrm{F}} = \mathrm{tr}^{1/2}(X^{\mathrm{T}}X)$ is the Frobenius norm of a matrix X, where $\mathrm{tr}(\cdot)$ represents the trace of a matrix.

2. Method and theoretical results

Our method is as follows. Let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be a sample generated from model (1). We divide $\mathcal{X}$ into two disjoint sets $\mathcal{X}_1 = \{X_1, \ldots, X_{[n/2]}\}$ and $\mathcal{X}_2 = \{X_{[n/2]+1}, \ldots, X_n\}$, where $[x]$ denotes the largest integer smaller than or equal to x. For a preset number $d > r$, let $Q_1 = \{(Q_{11}^{\mathrm{T}})_{p\times r}, (Q_{12}^{\mathrm{T}})_{p\times (d-r)}\}^{\mathrm{T}}$ and

$Q_2 = \{(Q_{21}^{\mathrm{T}})_{p\times r}, (Q_{22}^{\mathrm{T}})_{p\times(d-r)}\}^{\mathrm{T}}$ be $d \times p$ matrices of d eigenvectors of $\mathcal{X}_1\mathcal{X}_1^{\mathrm{T}}/(p[n/2])$ and $\mathcal{X}_2\mathcal{X}_2^{\mathrm{T}}/(p[n/2])$, respectively. The d eigenvectors in rows correspond to the d largest eigenvalues sorted in descending order. We perform sample splitting to ensure that Q_1 and $(\epsilon_{[n/2]+1}, \ldots, \epsilon_n)$ are weakly dependent and that Q_2 and $(\epsilon_1, \ldots, \epsilon_{[n/2]})$ are weakly dependent. This weak dependence diversifies the idiosyncratic component in the construction of principal components below. For example, when $\{(f_t^{\mathrm{T}}, \epsilon_t^{\mathrm{T}})^{\mathrm{T}}\}$ is serially independent, the matrix Q_{3-l} $(l = 1, 2)$ is independent of $\{\epsilon_{[n/2](l-1)+1}, \ldots, \epsilon_{[n/2](l-1)+[n/2]}\}$, and hence $\|Q_{3-l}\epsilon_{[n/2](l-1)+m}p^{-1/2}\| = O_{\mathrm{p}}(p^{-1/2})$ $(m = 1, \ldots, [n/2])$. This is akin to diversifying idiosyncratic risk in finance by constructing portfolios.

Now we have a series of d principal components or portfolios, $\tilde{X}_j = Q_2X_j$ and $\tilde{X}_{[n/2]+j} = Q_1X_{[n/2]+j}$ for $j = 1, \ldots, [n/2]$. Let $A_j = (\tilde{X}_{j+1}, \ldots, \tilde{X}_{j+d})p^{-1/2}$. In the Supplementary Material we show that

$$A_j \doteq \bar{A}_j = \left\{ \begin{array}{c} \bar{H}^{\mathrm{T}}(f_{j+1}, \ldots, f_{j+d}) \\ 0 \end{array} \right\}$$

for some matrix $\bar{H}$. This demonstrates that the cross-sectional averaging diversifies the idiosyncratic errors, and that estimating the number of factors is equivalent to estimating the common rank of the A_j asymptotically. To derive the theory, we need the following technical conditions.

Assumption 1. We have that $\Lambda^{\mathrm{T}}\Lambda/p = I_r$. The covariance matrix Σ_f of f_t is positive definite with distinct eigenvalues.

Assumption 2. For some $\theta > 0$, $\{(f_t^{\mathrm{T}}, \epsilon_t^{\mathrm{T}})^{\mathrm{T}}\}$ is an α-mixing stationary series satisfying

$$\alpha(k) = \sup_{A\in\mathcal{G}_{-\infty}^{0}, B\in\mathcal{G}_{k}^{\infty}} \left|\mathrm{pr}(A, B) - \mathrm{pr}(A)\,\mathrm{pr}(B)\right| \leqslant C\exp(-k^{\theta}),$$

where $\mathcal{G}_{-\infty}^{0}$ and $\mathcal{G}_{k}^{\infty}$ denote the σ-algebras generated by $\{f_t, \epsilon_t : t \leqslant 0\}$ and $\{f_t, \epsilon_t : t \geqslant k\}$, respectively, and $\max_{i\leqslant p, t\leqslant n} E\{|\epsilon_{it}|^{\max(2d,4)}\} + \max_{j\leqslant r, t\leqslant n} E\{|f_{jt}|^{\max(2d,4)}\} < \infty$.

Assumption 3. For vectors $(u_1, \ldots, u_p)^{\mathrm{T}}$ satisfying $\sum_{i=1}^{p} u_i^2 = 1$, we have that $\max_t E(\sum_{i=1}^{p} u_i\epsilon_{it})^4 < \infty$. For some constant C, we have that $E[p^{-1/2}\sum_{i=1}^{p}\{\epsilon_{it}\epsilon_{il} - E(\epsilon_{it}\epsilon_{il})\}]^2 < C$, $\|E\{e^{\mathrm{T}}e/(pn)\}\| \leqslant C/\min(p, n)$ where $e = (\epsilon_1, \ldots, \epsilon_n)$, and $E(\|\sum_{t=1}^{n} f_{jt}\epsilon_t\|^2) \leqslant Cn(p+n)$ for $j = 1, \ldots, r$.

Assumptions 1–3 are standard in the literature; see, for example, Chamberlain & Rothschild (1983), Onatski (2010) and Ahn & Horenstein (2013). Assumption 1 is a strong factor and partial identifiability condition. We assume strong factors to derive the distributional property of the rank estimator, but for consistency this assumption can be relaxed to the weak factor condition that $\Lambda^{\mathrm{T}}\Lambda/p^{\alpha}$ has bounded eigenvalues for some $0 < \alpha < 1$. Like the estimators in Bai & Ng (2002), Onatski (2010) and Ahn & Horenstein (2013), our rank estimator is consistent also for a class of time-varying factor loadings, but for mathematical neatness we assume that Λ is a constant matrix. Assumption 2 says that the joint series of factors and idiosyncratic errors are stationary and temporally weakly dependent. The exponential mixing rate implies that $E(\|Q_1\epsilon_{[n/2]+j}\|^2 + \|Q_2\epsilon_{[n/2]-j}\|^2) \leqslant C$ $(j = n^*, \ldots, [n/2])$ for some $n^* > 0$ satisfying $n^*n^{-1/2} \to 0$. Assumption 3 says that the idiosyncratic errors are cross-sectionally weakly dependent and also serially weakly dependent on the factors.

Let $Z_j = \mathrm{diag}\{z_{1j}, \ldots, z_{dj}\}$ be a diagonal matrix of independent and identically distributed random variables with zero mean and finite moment-generating functions. Inspired by the matrix expansion of Jacod & Podolskij (2013), we can prove that as $h \to 0$,

$$\det(A_j + hZ_j) = C_rh^{d-r} + o(h^{d-r}), \quad \det(A_j + \surd 2hZ_j) = C_r(\surd 2h)^{d-r} + o(h^{d-r})$$

for some random variable C_r having positive mean. This motivates us to estimate the rank of the A_j from $\hat{r}_h = d + \log{(S_{n,1}/S_{n,2})}/\log 2$, where

$$S_{n,1} = (n-d)^{-1}\sum_{j=0}^{n-d-1} \det^2(A_j + hZ_j), \quad S_{n,2} = (n-d)^{-1}\sum_{j=0}^{n-d-1} \det^2(A_j + \surd 2hZ_j).$$

We take squared determinants to guarantee a meaningful logarithmic function. Let $\gamma_{a_1,\ldots,a_m}(B_1,\ldots,B_m)$ be the sum of determinants over the class of matrices formed by choosing a_j $(j = 1,\ldots,m)$ columns from B_j, and let $A_{2j} = Q_{3-l}\{\epsilon_{[n/2](l-1)+j+1},\ldots,\epsilon_{[n/2](l-1)+j+d}\}/\surd p$ $(l = 1, 2)$.

Theorem 1. *Under Assumptions 1–3, if $p^{-1/2}h^{-1} = O(h)$ and $h = o(1)$, then*

$$\hat{r}_h - r = (\mu_0 \log 2)^{-1}\{z_{p,n,h} - \mu_1 h^2 + \mu_2/(2ph^2)\} + O_{\rm p}(p^{-1} + n^{-1} + h^4),$$

where $\mu_0 = E[\det^2\{\bar{H}^{\rm T}(f_{j+1},\ldots,f_{j+r})\}\prod_{m=1}^{d-r} Z_{r+m}^2]$, $\mu_1 = E\{\gamma_{r-1,d-r+1}^2(\bar{A}_j, Z_j)\}$, $\mu_2 = \sum_{j=0}^{n-d-1} pE\{\gamma_{r,1,d-r-1}^2$ $(\bar{A}_j, A_{2j}, Z_j)\}/(n-d)$, and

$$z_{p,n,h} = (n-d)^{-1}\sum_{j=0}^{n-d-1}\Big\{(1-\surd 2)h\gamma_{r,d-r}(\bar{A}_j, Z_j)\gamma_{r-1,d-r+1}(\bar{A}_j, Z_j) + (1-2^{-1/2})p^{-1/2}h^{-1}\gamma_{r,d-r}(\bar{A}_j, Z_j)\gamma_{r,1,d-r-1}(\bar{A}_j, \surd pA_{2j}, Z_j)\Big\}.$$

The rank estimator $\hat{r}_h$ has a continuous limit and assigns positive probabilities to intervals containing no integers, which seems meaningless since the factor number is certainly integer-valued. Theorem 1 demonstrates that this positive probability shrinks to zero as $p, n \to \infty$. Therefore, as in the proof of Theorem 3 of Trapani (2018), the interval probability of $\hat{r}_h$ can be interpreted as the mass probability assigned to integers within the interval. The expansion in Theorem 1 shows that the random-perturbation-based rank estimator has a bias of order $h^2 + (ph^2)^{-1}$ and a standard error term $z_{p,n,h}$ of order $(h + p^{-1/2}h^{-1})n^{-1/2}$, which is higher than the typical root-n convergence rate in the standard central limit theorem. To minimize the bias, an optimal choice of h is $cp^{-1/4}$ for some constant c. Theoretically, the optimal c is $c_0 = \{\mu_2/(2\mu_1)\}^{1/4}$, which makes the main bias term $\mu_2/(2ph^2) - \mu_1 h^2$ zero. Then the bias term of order h^4 is dominated by the standard error once $np^{-3/2} = o(1)$, which yields a central limit theorem of $\hat{r}_h$ as follows.

Corollary 1. *Under Assumptions 1–3, if $h = c_0 p^{-1/4}$ and $np^{-3/2} = o(1)$, then*

$$\nu_n^{-1}(n-d)^{1/2}c_0^{-1}p^{1/4}(\hat{r}_h - r) \to N(0, 1)$$

in distribution, where

$$\nu_n^2 = (n-d)^{-1}\sum_{j=0}^{n-d-1} E\Big\{(1-\surd 2)\gamma_{r,d-r}(\bar{A}_j, Z_j)\gamma_{r-1,d-r+1}(\bar{A}_j, Z_j) + (1-2^{-1/2})c_0^{-2}p^{1/2}\gamma_{r,d-r}(\bar{A}_j, Z_j)\gamma_{r,1,d-r+1}(\bar{A}_j, A_{2j}, Z_j)\Big\}^2\Big/(\mu_0 \log 2)^2.$$

To the best of our knowledge, no previous results exist for the convergence rate of the estimated number of factors. In an earlier unpublished work, this author proposed a less efficient rank estimator with a deterministic perturbation matrix hI_d, where I_d is the d-dimensional identity matrix. The principal bias term in that manuscript is $z_{p,n,h}$ with Z_j replaced by I_d, whose entries are not centred at zero, and hence $E(z_{p,n,h}) \neq 0$ when $\{f_t\}$ is serially correlated or when $\{f_t\}$ and $\{\epsilon_t\}$ are serially dependent. Therefore, its bias is of order $h + p^{-1/2}h^{-1}$, much lower than that of the estimator in the present paper. As a consequence, the deterministic-perturbation-based rank estimator converges at rate $n^{-1/2}$ under the stricter condition that $n/p^{1/2} = o(1)$. The increase in the convergence rate and relaxation of the condition for the new rank estimator is partially due to the random perturbation, which smooths the persistence of $\{f_t\}$ and the dependence between $\{f_t\}$ and $\{\epsilon_t\}$ contained in $z_{p,n,h}$. The other reason for the efficiency gain is the way in which the new estimator is constructed. In the Taylor expansion of $\log(S_{n,1}/S_{n,2})$, some of the major bias terms of $S_{n,1}$ and $S_{n,2}$ cancel out; see the Supplementary Material for details.

The parameter c controls the magnitude of the signal-to-perturbation ratio. When c is close to zero, the series of d principal components dominates the perturbation term, making $S_{n,1}$ and $S_{n,2}$ close to each other, and hence our estimate is close to d. When c is large, the perturbation term dominates and $S_{n,2}$ is approximately 2^d times $S_{n,1}$, so the estimate reduces to zero. This is clearly observed in the real-data analysis in the Supplementary Material. The bias term in Theorem 1 decreases and changes sign at c_0. In practice, we can make use of this monotonicity to find an appropriate c_0. For each c, we independently repeat the random perturbation and rank estimation M times. Denote the mth rank estimator and asymptotic error term by $\hat{r}_c(m)$ and $z_{p,n,c}(m)$, respectively. Let $\bar{r}_c = \sum_{m=1}^{M} \hat{r}_c(m)/M$. The expansion in Theorem 1 shows that

$$\bar{r}_c = r + \left\{ \mu_2/(2ph^2) - \mu_1 h^2 + \sum_{m=1}^{M} z_{p,n,c}(m)/M \right\} \Big/ (\mu_0 \log 2) + o_{\mathrm{p}}(hn^{-1/2}). \tag{2}$$

Independence implies that $\sum_{m=1}^{M} z_{p,n,c}(m)/M = O_{\mathrm{p}}\{h/(Mn)^{1/2}\} = o_{\mathrm{p}}(h/n^{1/2})$ as $M \to \infty$. A natural estimate $\hat{c}_0$ of c_0 is the solution to $\bar{r}_c = \mathrm{round}(\bar{r}_c)$, accounting for the fact that r must be an integer. Expansion (2) suggests an estimate $\hat{b}(c)$ of the bias term $b(c)$: $\hat{b}(c) = 2(\bar{r}_{\sqrt{2}h} + \bar{r}_{h/\sqrt{2}} - 2\bar{r}_h)$. It also implies that $\hat{b}(c) = b(c) + o_{\mathrm{p}}(hn^{-1/2})$ for c in a neighbourhood of c_0. Therefore, $\hat{c}_0$ is close to the root c^* of $\hat{b}(c) = 0$. Two steps lead to $\hat{c}_0$ in applications. First, find c^* and the integer $\mathrm{round}(\bar{r}_{c^*})$. Second, solve $\bar{r}_c = \mathrm{round}(\bar{r}_{c^*})$ and obtain $\hat{c}_0$. Then $\hat{c}_0 = c_0 + o_{\mathrm{p}}(h/\sqrt{n})$ as implied by (2).

To apply Corollary 1, one needs to estimate ν_n^2. In the Supplementary Material we show that $S_{n,1} h^{-2(d-\hat{r})} = \mu_0 + o_{\mathrm{p}}(1)$ and

$$\begin{aligned}\hat{r}_h - r = \{(n-d)\mu_0 \log 2\}^{-1} \sum_{j=0}^{n-d-1} \left\{ \det{}^2(A_j + hZ_j) - \det{}^2(A_j + \surd 2hZ_j) 2^{-(d-r)} \right\} h^{-2(d-r)} \\ + O_{\mathrm{p}}\{h^4 + \min(p,n)^{-1}\}.\end{aligned}$$

This motivates us to estimate ν_n^2 by

$$\begin{aligned}\hat{\nu}_n^2 = h^{-2}(n-d)^{-1} \sum_{j=0}^{n-d-1} \left[\left\{ \det{}^2(A_j + hZ_j) - \det{}^2(A_j + \surd 2hZ_j) 2^{-(d-\hat{r})} \right\} h^{-2(d-\hat{r})} \right]^2 \\ \times (S_{n,1} h^{-2(d-\hat{r})} \log 2)^{-2},\end{aligned}$$

where $\hat{r}$ is some consistent estimate of r, for example $\hat{r} = \mathrm{round}(\bar{r}_{c^*})$.

THEOREM 2. *Under Assumptions 1–3, if* $h = c_0 p^{-1/4}$ *and* $np^{-3/2} = o(1)$, *then* $\hat{\nu}_n^2 = \nu_n^2 + o_{\mathrm{p}}(1)$ *and* $t_{n,r} = \hat{\nu}_n^{-1} c_0^{-1} p^{1/4} (n-d)^{1/2} (\hat{r}_h - r) \to N(0,1)$ *in distribution.*

Theorem 2 demonstrates that we can test $H_0 : r = r_0$ versus $H_1 : r \mathrel{ǂ} r_0$, rejecting H_0 when $|t_{n,r_0}| > z_{\alpha/2}$ where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of $N(0,1)$. For a one-sided alternative $r < r_0$ or $r > r_0$, we reject H_0 when $t_{n,r_0} < -z_\alpha$ or $t_{n,r_0} > z_\alpha$. Theorem 2 also shows that the test has size close to the nominal level and power that is asymptotically equal to 1.

Table 1. *Estimates, standard errors and empirical sizes for testing* $H_0 : r = r_0$ *versus* $H_1 : r \neq r_0$ *with nominal levels* 10% *and* 5%

(p, n)		$r_0 = 1$	$r_0 = 2$	$r_0 = 3$	$r_0 = 4$	$r_0 = 5$	$r_0 = 6$
(150, 200)	e (se)	1.05 (0.09)	2.02 (0.12)	3.02 (0.15)	4.00 (0.16)	5.00 (0.16)	5.99 (0.15)
	(10%, 5%)	(0.15, 0.08)	(0.12, 0.05)	(0.12, 0.06)	(0.10, 0.05)	(0.12, 0.05)	(0.11, 0.05)
(300, 200)	e (se)	1.02 (0.07)	2.00 (0.10)	3.02 (0.12)	4.00 (0.13)	5.01 (0.13)	6.00 (0.13)
	(10%, 5%)	(0.12, 0.06)	(0.11, 0.05)	(0.12, 0.05)	(0.10, 0.05)	(0.10, 0.05)	(0.12, 0.06)
(240, 360)	e (se)	1.04 (0.06)	2.01 (0.08)	3.02 (0.10)	4.01 (0.11)	5.01 (0.11)	6.01 (0.11)
	(10%, 5%)	(0.14, 0.08)	(0.13, 0.07)	(0.11, 0.06)	(0.10, 0.05)	(0.11, 0.05)	(0.12, 0.05)
(360, 360)	e (se)	1.01 (0.05)	2.00 (0.07)	3.02 (0.09)	4.01 (0.10)	5.01 (0.10)	6.01 (0.10)
	(10%, 5%)	(0.11, 0.05)	(0.11, 0.06)	(0.10, 0.05)	(0.11, 0.05)	(0.10, 0.04)	(0.11, 0.06)
(540, 360)	e (se)	1.01 (0.04)	2.00 (0.06)	3.01 (0.07)	4.01 (0.09)	5.01 (0.09)	6.01 (0.09)
	(10%, 5%)	(0.11, 0.06)	(0.11, 0.05)	(0.11, 0.04)	(0.09, 0.04)	(0.10, 0.05)	(0.11, 0.04)
(360, 500)	e (se)	1.02 (0.04)	2.01 (0.06)	3.02 (0.07)	4.01 (0.08)	5.01 (0.09)	6.01 (0.08)
	(10%, 5%)	(0.14, 0.07)	(0.11, 0.05)	(0.11, 0.06)	(0.10, 0.05)	(0.11, 0.05)	(0.10, 0.06)
(540, 500)	e (se)	1.01 (0.04)	2.00 (0.05)	3.01 (0.07)	4.01 (0.07)	5.02 (0.08)	6.01 (0.08)
	(10%, 5%)	(0.11, 0.06)	(0.11, 0.06)	(0.11, 0.05)	(0.10, 0.05)	(0.11, 0.05)	(0.10, 0.05)

e, estimate; se, standard error.

Table 2. *Empirical power and comparison with the Onatski* (2009) *test for* $H_0 : r = 3$ *versus the alternatives* $H_1 : r < 3$, $H_2 : r \neq 3$ *and* $H_3 : r > 3$

	r	1	2	3	4	5	6	7	8
H_1 (5%, 2.5%)	K	(1, 1)	(1, 1)	(0.059, 0.035)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
	ON	—	—	—	—	—	—	—	—
H_2 (10%, 5%)	K	(1, 1)	(1, 1)	(1, 1)	(1, 1)	(1, 1)	(1, 1)	(1, 1)	(1, 1)
	ON	—	—	—	—	—	—	—	—
H_3 (5%, 2.5%)	K	(0, 0)	(0, 0)	(0.056, 0.023)	(1, 1)	(1, 1)	(1, 1)	(1, 1)	(1, 1)
	ON	—	—	(0.065, 0.030)	(1, 1)	(1, 1)	(1, 1)	(1, 1)	(1, 1)

K, the test proposed in this paper; ON, the test of Onatski (2009).

3. Numerical experiment

We generate data from model (1) with $f_t = 0.1 f_{t-1} + 0.2 f_{t-2} + \epsilon_t^*$, where ϵ_t^* is a vector of standard normal random variables and $\lambda_{it} = \surd p\, I\{(t-1)p/r < i \leqslant tp/r\}$. As in Ahn & Horenstein (2013) and Bai & Ng (2002), we generate the idiosyncratic errors from $\epsilon_{it} = \rho \epsilon_{i(t-1)} + v_{it} + \sum_{j\neq 0, j=-J}^{J} \beta v_{(i-j)t}$ where the v_{it} $(i = 1, \ldots, p;\ t = 1, \ldots, n)$ are independently generated from standard normal distributions. The parameters ρ and β control the temporal and cross-sectional correlations, respectively. We set $\rho = 0.3$, $\beta = 0.1$, $J = 1$ and $d = 8$. We use one round of simulation to determine c_0 and then fix it in subsequent repetitions. For example, when $r = 2$, the estimate $\hat{c}_0$ is always around 1.95 for all (n, p) in our search for $c \in [1.375, 2.375]$ with $M = 2000$. The entries of Z_j take values -1 and 1 with equal probability.

The goal is to assess the performance of our estimates in hypothesis testing in cases where existing estimates fail to give a testing rule. Table 1 displays the point estimates $\hat{r}_h$, standard errors and empirical sizes for testing $H_0 : r = r_0$ versus $H_1 : r \neq r_0$. It can be seen that our estimates are accurate and that the test performs well in terms of controlling Type I errors. Next, we evaluate the power performance of our test and compare it with the popular test of Onatski (2009). We report only the results for $p = 300$, $n = 200$ and $r_0 = 3$; the results for other cases are similar. Table 2 shows that both our test and that of Onatski (2009) are powerful and can test right-sided alternatives, but the Onatski (2009) test is not applicable to left-sided and two-sided alternatives while our test is.

Acknowledgement

This work was supported in part by the Natural Science Foundation of China and PAPD of Jiangsu Higher Education Institutions. The author thanks the editor, associate editor and reviewers for helpful comments.

Supplementary material

Supplementary material available at *Biometrika* online includes additional simulation studies, a real-data analysis, technical lemmas and propositions with full proofs, and proofs of the main theorems.

References

Ahn, S. & Horenstein, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–27.

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

Chamberlain, G. & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51**, 1281–304.

Connor, G. & Korajczyk, R. (1993). A test for the number of factors in an approximate factor model. *J. Finan.* **48**, 1263–91.

Cragg, J. & Donald, S. (1997). Inferring the rank of a matrix. *J. Economet.* **76**, 223–50.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Statist.* **82**, 540–54.

Hallin, M. & Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *J. Am. Statist. Assoc.* **102**, 603–17.

Jacod, J. & Podolskij, M. (2013). A test for rank of the volatility process: The random perturbation approach. *Ann. Statist.* **41**, 2391–427.

Kleibergen, F. & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *J. Economet.* **133**, 97–126.

Kong, X. (2017). On the number of common factors with high-frequency data. *Biometrika* **104**, 397–410.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* **77**, 1447–79.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Statist.* **92**, 1004–16.

Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. *J. Am. Statist. Assoc.* **113**, 1341–49.

[*Received on* 20 *November* 2018. *Editorial decision on* 22 *August* 2019]