Determining the number of factors for high-dimensional time series

Article *in* Statistics and its interface · March 2018

DOI: 10.4310/SII.2018.v11.n2.a8

Qiang Xia
South China Agricultural University

Jianhong Wu
Shanghai Normal University

Heung Wong
The Hong Kong Polytechnic University

Determining the number of factors for high-dimensional time series

Qiang Xia[a] Rubing Liang[a*] Jianhong Wu[b] Heung Wong[c]

[a] *College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China*

[b] *School of Mathematics and Science, Shanghai Normal University, Shanghai 200234, China*

[c] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong*

Abstract. In this paper, we suggest a new method of determining the number of factors in factor modeling for high-dimensional stationary time series. When the factors are of different degree of strength, the eigenvalue-based ratio method of Lam and Yao needs a two-step procedure to estimate the number of factors. As a modification of the method, however, our method only needs a one-step procedure for the determination of the number of factors. The resulted estimator is obtained simply by minimizing the ratio of the contribution of two adjacent eigenvalues. Some asymptotic results are also developed for the proposed method. The finite sample performance of the method is well examined and compared with some competitors in the existing literature by Monte Carlo simulations and a real data analysis.

JEL Classification:

Keywords: Autocovariance matrices; Contribution ratio; Eigenvalues; Factor models; Number of factors

*Corresponding author: Rubing Liang, College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642. Email: stat2003@163.com.

1 Introduction

The availability of large or vast time series data brings the opportunities as well as challenges to time series analysts. More and more people pay attention to the analysis of high-dimensional time series, which will be of increased interest and importance in the modern information age. For example, it is the key for the financial market analysts to understand the dynamics of the returns of a large number of assets. The study of economic and business phenomena encounters both large numbers of cross-section units and time series observations frequently. Due to the large number of indices monitored across many different locations, environmental time series are often high dimensional. As argued by Lam and Yao (2012), the standard multiple time series models such as vector ARMA models are seldom used directly in practice due to the problem of overparametrization. More and more people focus on factor modelling which can provide a low-dimensional and parsimonious representation for high-dimensional dynamics. In the factor modelling for economic or financial data with both high dimension (d) and time series observations (n), one of the fundamental issues is how to determine the number of common factors. Recent attempts in this direction included Bai and Ng (2002), Onatski (2010), Alessi, Barigozzi, and Capasso (2010), Ahn and Horenstein (2013), Wu (2016), and Xia, Liang, and Wu (2017), which mainly focused on the static factor model. Forni et al. (2000), Hallin and Liska (2007), Amengual and Watson (2007) and Onatski (2009), worked on the dynamic factor model. Pan and Yao (2008), Lam and Yao (2012), Xia, Xu, and Zhu (2015) and Chan, Lu, and Yau (2016), focused on determining a few common factors for high-dimensional time series.

Although the above literatures have considered statistical methodologies to determine the number of factors, their estimators are not directly comparable as they require different restrictive conditions. For example, in the frequency domain, Forni et al. (2000) considered an information criterion to estimate the number of factors based on the portion of explained variances in the dynamic factor model. In the time domain, Bai and Ng (2002) proposed the information criteria of model selection named panel and information C_p (PC and IC) criteria for determining the number of factors in approximate factor models, but the weak cross-sectional dependence on the idiosyncratic component was required. Relied on the information from the autocovariance matrices at nonzero lag, Lam and Yao (2012) adopted an "Eigenvalue Ratio" (ER) approach to determine the number of factors for high-dimensional time series. In a different setting, based on the variance-covariance matrix, Ahn and Horenstein (2013) proposed the ER estimator as well as "Growth Ratio" (GR) method to estimate the number factors in approximate static factor models. In this paper, we follow the framework of Lam and Yao (2012), which enjoys the advantages that the idiosyncratic component can have strong cross-sectional dependence and it allows auto-correlation and cross-correlation between factors and idiosyncratic component. In particular, the ER estimator for the number of factors by Lam and Yao (2012) has made contributions to the associated asymptotic theory which underpins the "blessing of dimensionality" phenomenon observed in numerical experiments. Lam and Yao (2012) also suggested some further problems about estimating the number of factors for high-dimensional time series.

To our knowledge, the ER estimator is not consistent. Xia et al. (2015) suggested a ridge-type ratio estimator (RER) to solve this problem by modifying the ER estimate. The performance of the RER estimator, however, depends on the selection of a positive value c. There is yet no simple method to determine the value of c. Moreover, a fly in the ointment of the eigenvalue-ratio-based estimators is the need of a two-step estimation procedure to determine the number of factors, even when the factors are of different degrees of strength. In practice, a problem is whether one should use two steps to estimate the number of factors. In general, one-step procedure is more desirable for estimating the number of factors. Meanwhile, given the importance of the problem, it is meaningful to develop a new method which only needs one step to estimate the number of factors in factor models for high-dimensional time series. Based on the ratio of two adjacent eigenvalues, we can consider each of its proportional value in a set of relative eigenvalues, which can be regarded as the contribution of the eigenvalue. Thus, we suggest a new estimator named as "contribution ratio" (CR) of two adjacent eigenvalues. See equation (2.6) and its explanation for the meaning of CR.

It can be further shown that, comparing with the competitors in the existing literature, the CR estimator has desired performance on determining the number of common factors, especially when both strong and weak factors exist in the factor models. Through Monte Carlo simulation experiments and a real data application, we find that this new method can improve the performance of estimation for the number of factors, meanwhile, when the factors are of different degrees of strength in many cases, we can determine the number of factors by one step.

The rest of the paper is organized as follows. Section 2 introduces the methodology based on eigenanalysis for an autocovariance matrix. Asymptotic properties of the proposed estimator are investigated in Section 3. Both simulation study and a real data analysis are presented in Section 4. All technical arguments are relegated to the Appendix.

Throughout this paper, A' means the transpose of the matrix A, and $\| M \|_{max}$ and $\| M \|_{min}$ the positive square root of the maximum and minimum eigenvalue of MM', respectively. If $a = O_p(b)$ and $b = O_p(a)$, then we denote $a \asymp b$.

2 Models and Estimation Methodology

2.1 Factor Models

For $t = 1, ..., n$, let Y_t be a d-dimensional multivariate time series. As argued by Lam and Yao (2012), Y_t can include two parts: a common component X_t driven by a low-dimensional process and a static idiosyncratic component ξ_t which is a white noise process. We then have the decomposition as follows,

$$Y_t = AX_t + \xi_t, \tag{2.1}$$

where $X_t = (x_{1t}, ..., x_{rt})'$ is the r-dimensional vector of unobserved common factors, $r < d$ is unknown; A is a $d \times r$ factor loading matrix; $\{\xi_t\}$ is a vector white-noise process, i.e., $\xi_t \sim WN(\mu_\xi, \Sigma_\xi)$. It is worthy to point out that the model (2.1) is not identified because for any $r \times r$ nonsingular matrix H the observed series Y_t can be expressed in terms of a new set of factors, i.e.,

$$Y_t = AHH^{-1}X_t + \xi_t. \tag{2.2}$$

However, the linear space spanned by the columns of A, denoted by $\mathcal{M}(A)$ and called as the factor loading space, is uniquely defined by (2.1), i.e., $\mathcal{M}(A) = \mathcal{M}(AH)$. In this sense the factor loading matrix A is unique, once it is specified, then the factor process X_t is also uniquely defined accordingly. In this paper, we choose a half orthogonal matrix A, with $A'A = I_r$, where I_r is the $r \times r$ identity matrix. Therefore, we can always rotate an estimated factor loading matrix whenever appropriate, this helps us facilitate our estimation for A in a simple and convenient manner.

2.2 Estimation of r

To obtain our asymptotic theory, we introduce some regularity assumptions as follows. Denote $\Sigma_Y(l) = cov(Y_{t+l}, Y_t), \Sigma_X(l) = cov(X_{t+l}, X_t)$, $\Sigma_{\xi X}(l) = cov(\xi_{t+l}, X_t)$ and $\Sigma_{X\xi}(l) = cov(X_{t+l}, \xi_t)$, for all $l \geq 1$.

Assumption 1. An arbitrary linear combination of the components of X_t is not white noise, and $A'A = I_r$.

Assumption 2. For $l = 0, 1, ..., l_0$, where $l_0 \geq 1$, and a constant $\delta \in [0, 1]$, $\Sigma_X(l)$ is of full rank,

$$\parallel \Sigma_X(l) \parallel_{max} \asymp d^{1-\delta} \asymp \parallel \Sigma_X(l) \parallel_{min} .$$

Assumption 3. For $l \geq 0$, $\Sigma_{X\xi}(l)$ and Σ_ξ remains bounded elementwisely, as d and n increase to infinity. Also, $\parallel \Sigma_{X\xi}(l) \parallel_2 = o(d^{1-\delta})$.

Assumption 4. The covariance matrix $\Sigma_{\xi X}(l) = 0$ for all $l > 0$.

Assumption 5. The time series$\{Y_t\}$ is strictly stationary and ψ-mixing, where the mixing coefficients $\psi(\cdot)$ satisfy the condition $\Sigma_{t\geq 1} t\psi(t) < \infty$. Furthermore, $E(\mid Y_t \mid^4) < \infty$.

Remark 1. Here δ is taken as a measure of the strength of the factor in Assumption 2. When $\delta = 0$, the corresponding factors are called strong factors, and when $\delta > 0$ the factors are called weak factors, which is different from Onatski (2012). The detailed description of factor strength can be found in Part 3.2 of Lam and Yao (2012). Assumption 3 requires that the correlation between X_{t+l} and ε_t is not too strong. Assumption 4 relaxes the traditional independent assumption between factor process and noise process.

According to Assumption 4, we have $\Sigma_Y(l) = A\Sigma_X(l)A' + A\Sigma_{X\xi}(l)$. Let

$$\Omega = \sum_{l=1}^{l_0} \frac{\Sigma_Y(l)\Sigma_Y(l)'}{d^2},$$

where l_0 is a prescribed positive integer. Since $\Sigma_Y(0) = A\Sigma_X(0)A' + \Sigma_\xi(0)$ and $\Sigma_Y(0)B \neq 0$, then $l = 0$ is excluded from the sum in Ω.

The matrix Ω has the two good important properties as demonstrated by Lam et al. (2011) and Lam and Yao (2012). On one hand, it is a nonnegative definite matrix, of which the eigenvalues are real and nonnegative. On the other hand, if $\Omega B = 0$, then $\Sigma'_Y(l)B = 0$ for all $l \geq 1$. Based on these two aspects, the eigenvectors of Ω corresponding to different eigenvalues are orthogonal to each other. Therefore, the number of factors r is the number of non-zero eigenvalues of Ω. And then the r orthonormal eigenvectors of Ω corresponding to its non-zero eigenvalues are the columns of A. Hence, to estimate both r and A, it is needed to perform an eigenanalysis for the sample counterpart of Ω is needed as follows,

$$\hat{\Omega} = \sum_{l=1}^{l_0} \frac{S_l S_l'}{d^2}, \tag{2.3}$$

where $S_l = \dfrac{1}{n-l} \sum_{t=1}^{n-l} (Y_{t+l} - \bar{Y})(Y_t - \bar{Y})'$ and $\bar{Y} = \dfrac{1}{n} \sum_{t=1}^{n} Y_t$.

Assumption 6. Let $\lambda_1, ..., \lambda_d$ be the eigenvalues of the matrix Ω, then

$$\lambda_1 > \lambda_2 > ... > \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_d.$$

Remark 2. The Assumptions 1-6 are the same as those of Lam and Yao (2012).

Since the estimates for the zero-eigenvalues of Ω are not exactly the zero values in a finite sample, the first nonzero-eigenvalues can not be used for the estimation of r directly. Some available methods for determining r can be found in the existing literature. For example, Lam and Yao (2012) suggested the ER estimator for r. They plotted all the estimated eigenvalues in a descending order, and looked for a cutoff $\hat{r}$ such that the $\hat{r}$th largest eigenvalue is substantially larger than the $(\hat{r}+1)$th largest eigenvalues. This ER estimator defined below benefits from the faster convergence rates of the estimators for the zero-eigenvalues, and may be viewed as an enhanced eyeball test.

$$\hat{r}_{ER} = arg \min_{1 \leq i \leq R} \frac{\hat{\lambda}_{i+1}}{\hat{\lambda}_i} =: arg \min_{1 \leq i \leq R} ER(i), \tag{2.4}$$

where $\hat{\lambda}_1 \geq ... \geq \hat{\lambda}_d$ are the eigenvalues of $\hat{\Omega}$, and $r < R < d$ is a constant.

Based on the variance-covariance matrix, Ahn and Horenstein (2013) also consider ER estimator by putting (2.4) upside down. To estimate the number factors, they use the GR estimator, which is given by

$$\hat{r}_{GR} = \underset{1\le i\le R}{argmin} \frac{ln[\sum\limits_{k=i+1}^{m} \hat{\lambda}_k / \sum\limits_{k=i+2}^{m} \hat{\lambda}_k]}{ln[\sum\limits_{k=i}^{m} \hat{\lambda}_k / \sum\limits_{k=i+1}^{m} \hat{\lambda}_k]} =: \underset{1\le i\le R}{argmin} GR(i), \tag{2.5}$$

where $m = min\{d, n\}$. The GR estimator can also be used for the factor model of Lam and Yao (2012).

Different from the above ER and GR estimators, we consider a new criterion function. Because $\hat{\lambda}_i / \sum\limits_{k=i}^{m} \hat{\lambda}_k$ denotes the contribution of $\hat{\lambda}_i$ for $\sum\limits_{k=i}^{m} \hat{\lambda}_k$, a contribution ratio (hereafter CR) is defined as follows,

$$CR(i) = \frac{\hat{\lambda}_{i+1} / \sum\limits_{k=i+1}^{m} \hat{\lambda}_k}{\hat{\lambda}_i / \sum\limits_{k=i}^{m} \hat{\lambda}_k},$$

and then we can obtain the estimator

$$\hat{r}_{CR} = arg \min_{1\le i\le R} CR(i), \tag{2.6}$$

where $m = min\{d, n\}$.

Remark 3. Note that the CR estimator is about the ratio of the two adjacent contribution ratios, which also add more useful information based on the ER estimator. In particular, when both a weak and a strong factors exist, the first eigenvalue is too large to enable ER(1) to be the smallest. Compare to the ER estimator, the relative contribution can weaken the dominant position of the first eigenvalue. This is due to the fact that

$$CR(i) = \frac{\hat{\lambda}_{i+1} / \sum\limits_{k=i+1}^{m} \hat{\lambda}_k}{\hat{\lambda}_i / \sum\limits_{k=i}^{m} \hat{\lambda}_k} = \frac{\hat{\lambda}_{i+1}}{\hat{\lambda}_i} + \frac{\hat{\lambda}_{i+1}}{\sum\limits_{k=i+1}^{m} \hat{\lambda}_k} = ER(i) + \frac{\hat{\lambda}_{i+1}}{\sum\limits_{k=i+1}^{m} \hat{\lambda}_k},$$

which also considers the contribution of $\hat{\lambda}_{i+1}$ based on the ER estimator at the same time, which means CR method can be effective in the search for the useful eigenvalues. Therefore, it can improve the performance of estimating the number of factors.

Remark 4. According to our experience in the simulations, as $d > n$, the finite sample performance of ER and GR approaches are sensitive to the choice of the possible maximum number of factors (R), but the CR estimator is rather robust to it because of its structure. In fact, as long as the last few eigenvalues are excluded, bad influence can be avoided. Hence, $R = min\{d, n\}/2$ is a good choice for ER, GR and CR approaches in practice.

Theorem 1. *Under the Assumptions 1-6, as $n \to \infty$, $d \to \infty$, $n = O(d^{2\delta+1})$ and $h_n = d^{\delta}n^{-1/2} \to 0$, then $CR(i) \asymp 1$, for $i = 1, ..., r-1$, and $CR(r) = O_P(d^{2\delta}n^{-1}) \xrightarrow{P} 0$.*

Corollary 1. *Under the conditions of Theorem 1, then*

$$CR(i) \asymp GR(i) \asymp ER(i), \quad for \quad i = 1, ..., r.$$

Remark 5. By Theorem 1, $h_n = d^{\delta}n^{-1/2} \to 0$ is needed, thus, if $d \asymp n$, then $\delta < 0.5$.

To improve the rates for the estimated eigenvalues, Lam and Yao (2012) entertained some additional conditions on their factor model. Under these conditions, we can discuss the asymptotic results of the CR estimator.

Assumption 7. Let ξ_{jt} denote the jth component of ξ_t. Then ξ_{jt} are independent for different t and j, and have mean 0 and common variance $\sigma^2 < \infty$.

Assumption 8. The distribution of each ξ_{jt} is symmetric. Furthermore, $E(\xi_{jt}^{2k+1}) = 0$, and $E(\xi_{jt}^{2k}) \le (\tau k)^k$ for all $1 \le j \le d$ and $t, k \ge 1$, where $\tau > 0$ is a constant independent of j, t, k.

Assumption 9. All the eigenvalues of Σ_ξ are uniformly bounded as $d \to \infty$.

Theorem 2. *Under the Assumptions 1-8, $\varrho_n = d^{\delta/2}n^{-1/2} \to 0$ and $n = O(d)$. Then, as $n \to \infty$, $d \to \infty$,*

$$CR(r) = O_P(d^{\delta}n^{-1}) \xrightarrow{P} 0.$$

If in addition Assumption 9 holds, we have

$$CR(r) = O_P(d^{-1/2+\delta}n^{-1} + m^{-1}).$$

Corollary 2. *Under the Assumptions 1-8, as $n \to \infty$, $d \to \infty$ and $n = O(d)$, then*

$$CR(r) \asymp GR(r) \asymp ER(r).$$

Remark 6. Comparing with Theorem 1, as $d \asymp n$ and all factors are weak ($\delta \neq 0$), then the speed at which CR(r) converges to zero increases. However, when all factors are strong, i.e., $\delta = 0$, the speed at which CR(r) converges to zero is unchangeable for d under Assumptions 1-8. But under Assumptions 1-9, the speed at which CR(r) converges to zero increases again whether $\delta = 0$ or not.

3 Simulation and Application

In this section, to investigate the overall performance of the CR estimator, we carry out five simulation experiments and a study based on one real data. Since the choice of l_0 is not sensitive to the estimate of r (Lam and Yao, 2012), we set $l_0 = 1$ for all simulations. The CR estimator will be used to analyse a real data set.

Meanwhile, the IC1 estimator of Bai and Ng (2002), the ED estimator of Onatski (2010) and the GR estimator of Ahn and Horenstein (2013) are also used to estimate the number of factors for this high-dimensional time series factor model (2.1). The performances of these estimators are compared to our CR estimator, the ER estimator of Lam and Yao (2012), and the RER estimator of Xia et al. (2015).

Remark 7. In fact, we can select $R = min\{d, n\}/2$ in the simulation experiments and the real data analysis for ER, RER, GR and CR estimators. However, the estimators of Bai and Ng (2002) and Onatski (2010) are sensitive to the choice of R. Small R is appropriate to them. Hence, we include $R = 5$ for the case with $n = 50$, $d = 10$, and $R = 10$ for the case with $n = 100$, $d = 20$. Other than these two cases, we select $R = 20$ in the other cases in all simulations, which is similar to Chan et al. (2016).

3.1 Simulation Experiments

To highlight the asymptotic properties in the previous section, some simulation examples, which are similar to Part (3.3) of Lam and Yao (2012), are conducted. We set in model (2.1) $r = 3$, $n = 50, 100, 200, 500$ and 1000, and $d = 0.2n, 0.5n, 0.8n$ and $1.2n$. All the $d \times r$ elements of A

are generated independently from the uniform distribution on the interval $[-1, 1]$ first, and then three factors, with each measure of the strength δ_i, are made by dividing each of them by $d^{\delta_i/2}$, respectively. We generate factor x_t from a 3×1 vector-AR(1) process with independent $N(0, 1)$ innovations and the diagonal autoregressive coefficient matrix with 0.8, -0.5 and 0.3 as the main diagonal elements. We let $\{\xi_t\}$ in model (2.1) consist of independent $N(0, 1)$ components and they are also independent across t. For each setting, we replicate the simulation 200 times for the following five cases individually.

(i) Case 1: $\delta_1 = \delta_2 = \delta_3 = 0$;

(ii) Case 2: $\delta_1 = \delta_2 = \delta_3 = 0.3$;

(iii) Case 3: $\delta_1 = \delta_2 = 0$, $\delta_3 = 0.3$;

(iv) Case 4: $\delta_1 = 0$, $\delta_2 = \delta_3 = 0.3$;

(v) Case 5: $\delta_1 = 0$, $\delta_2 = 0.4$, $\delta_3 = 0.2$.

The respective frequency estimates (x) for the probability $P(\hat{r} = 3)$, the number of underestimation (y) and overestimation (z) in the 200 replications are recorded by $x(y|z)$ in Tables 1∼5. That is to say that x is the frequency estimate for correction identification, and $200 - y - z$ is the correct number of estimation of different estimators.

When the factors are all strong, from Table 1, it seems that the CR and GR estimates are slightly better than the other estimators with $n \leq 100$. The CR estimate is as good as the ER, RER and GR estimates with $n \geq 200$. In this case, the IC1 and ED estimators do not perform well, they seem to encounter "the curse of dimensionality". Also, the IC1 and ED estimators often overestimate the number of factors, the peak of estimation for the ED estimator is at 4.

When the factors are all weak, Table 2 indicates that IC1 estimator performs better than the other estimators with $n = 50$, while the CR estimator works equally well or better than the other estimators with $n \geq 100$. The RCR and ED estimators are not performing satisfactorily, and they face easily "the curse of dimensionality" in Case 2.

In Case 3 and Case 4, there are two strong factors and one weak factor (or one strong factor and two weak factors) in the factor model. From Table 3 and Table 4, we can see that the ER and RER estimators are not effective. The peak of their estimations is at 1 (or 2). The ED estimator is always effective with small d $(= 0.2n)$. The IC1 estimator performs better than the other estimators with $n = 50$, while the CR estimator outperforms the others with $n = 100$ and 200. When $n \geq 500$, the CR and GR estimators perform better than the other estimators. The ED estimator still runs into "the curse of dimensionality" easily in these two cases.

When the strengthes of all factors are different in Case 5, Table 5 tells us that the CR estimator outperforms the other estimators with $n \leq 200$. When $n \geq 500$, the performance of GR estimator is as good as the CR estimator, which performs better than the other estimators.

Through these simulation results, the CR estimator has an advantage in determining the number of factors over the ER, RER, IC1, ED and GR estimators. In particular, when both weak factors and strong factors are in the factor model, the CR estimator only need to utilize the one-step procedure for determining the number of factors. The ER and RER estimators tend to underestimate the number of factors, while the IC1 and ED estimators tend to overestimate. Also, when both weak and strong factors exist in the factor model, the ER and RER estimators tend to identify the strong factors by the one-step procedure, and the ED estimator is prone to encounter "the curse of dimensionality".

3.2 A Real Data Example

We study the real data set about the daily returns of 123 stocks in the period 2 January 2002 to 11 July 2008, which was firstly analyzed by Lam and Yao (2012). Those stocks were selected among those included in the S&P500 and were traded every day during the period. The log-returns were multiplied by 100 based on the daily close prices. We have in total $n = 1642$ observations with $d = 123$. Firstly, we apply the eigenanalysis to the matrix $\hat{\Omega}$ defined in (2.3) with $l_0 = 1$. The relevant ER, RER, IC1, ED, GR and CR estimators' values with $R = 20$ are plotted in Figure 1. From which, we can see the estimated results of ER, RER, IC1, ED, GR and CR estimators. It is clear that the ER and RER estimators indicate $\hat{r} = 1$, the IC1 estimator shows $\hat{r} = 7$, while the ED, GR and CR estimators suggest $\hat{r} = 3$. According to the simulations, we believe that the final estimate should be 3 for the number of factors.

Meanwhile, we will check whether the estimation results are sensitive to the choice of R.

For $R = 60, 100$, the ER and RER estimators always gives $\hat{r} = 1$, while the IC1 estimator always gives $\hat{r} = 7$. But the ED estimator is easily influenced by R, because it finds 4 factors when $R = 60$; 5 factors when $R = 100$. Although the GR and CR estimators still give $\hat{r} = 3$, comparing with the CR estimator, the discriminations between the second and third positions of the ratio are not obvious for the GR estimator. This can be seen from Figure 1 with $R = 20$, which is similar to the case of $R = 60, 100$.

4 Conclusion

In this paper, based on the information of the autocovariance matrices at nonzero lags, we have introduced a new CR estimator to determine the number of common factors for the factor model of Lam and Yao (2012). Through some simulation experiments conducted to compare with ER estimator of Lam and Yao (2012), RER estimator of Xia et al. (2015), IC1 estimator of Bai and Ng (2002), ED estimator of Onatski (2010) and GR estimator of Ahn and Horenstein (2013), we found the CR estimator is easy to compute and can improve the performance of correctly estimating the number of factors for high-dimensional time series. When factors in the factor model are all strong or weak, the simulation results indicate that the CR estimator can perform equally to or better than the other estimators. The CR estimator, however, outperforms the competing estimators when the factors are of different degrees of strength.

Acknowledgments

The research of Rubing Liang was supported in part by National Science Foundation of Guangdong Province of China (No.2015A030310365) and Ministry of Education in China Project of Humanities and Social Sciences (No.17TJ3). The research of Qiang Xia was supported in part by National Science Foundation of Guangdong Province of China (No.2016A030313414) and National statistical plan for scientific research project of China (No.2015LZ48). The research of Jianhong Wu was supported in part by National Science Foundation of China (No.11671263). The research of Heung Wong was supported by a grant from the Research Committee of The Hong Kong Polytechnic University (G-YBCV).

Appendix: Proofs

Lemma 1. Under the Assumptions 1-6, and $h_n = d^{\delta} n^{-1/2} \to 0$. As $n \to \infty$ and $d \to \infty$, then

(i) $\mid \hat{\lambda}_i - \lambda_i \mid = O_P(d^{-\delta} n^{-1/2})$ for $i = 1, ..., r$,

(ii) $\mid \hat{\lambda}_i \mid = O_P(n^{-1})$ for $i = r+1, ..., d$, and

(iii) $ER(i) \asymp 1$ for $i = 1, ..., r-1$, and $ER(r) = O_P(d^{2\delta} n^{-1})$.

Proof of Lemma 1. Firstly, some elementary calculations lead to the following decomposition:

$$\| \hat{\Omega} - \Omega \| \leq \frac{1}{d^2} \sum_{l=1}^{l_0} \left[\| S_l - \Sigma_Y(l) \|^2 + 2 \| \Sigma_Y(l) \| \| S_l - \Sigma_Y(l) \| \right].$$

Since A is assumed to be a half orthogonal matrix in model (2.1), by Assumptions 2 and 3, we have

$$\| \Sigma_Y(l) \| = \| A\Sigma_X(l)A' + A\Sigma_{X\xi}(l) \| \leq \| \Sigma_X(l) \| + \| \Sigma_{X\xi}(l) \| = d^{1-\delta}.$$

Also, by Lemma 2 of Lam et al. (2011), we have

$$\| S_l - \Sigma_Y(l) \| \leq \| \hat{\Sigma}_X(l) - \Sigma_X(l) \| + 2 \| \hat{\Sigma}_{X\xi}(l) - \Sigma_{X\xi}(l) \| + \| \hat{\Sigma}_{\xi}(l) \| = O_p(dn^{-1/2}),$$

where $\hat{\Sigma}_X(l)$ and $\hat{\Sigma}_{X\xi}(l)$ are the sample variances of $\Sigma_X(l)$ and $\Sigma_{X\xi}(l)$, and $\| \hat{\Sigma}_{\xi}(l) \| \leq \| \hat{\Sigma}_{\xi}(l) \|_F$ $= O_p(dn^{-1/2})$ with $\| \hat{\Sigma}_{\xi}(l) \|_F = trace(\hat{\Sigma}_{\xi}(l)\hat{\Sigma}'_{\xi}(l))$. Then, we have

$$\| \hat{\Omega} - \Omega \| = O_p(d^{-\delta} n^{-1/2}).$$

Moreover, it holds that $A^{\tau}\Omega A = D$, where D is diagonal with the presentation:

$$D = \frac{1}{d^2} \sum_{l=1}^{l_0} \left[\Sigma_X(l)A' + \Sigma_{X\xi}(l)\right] \left[\Sigma_X(l)A' + A\Sigma_{X\xi}(l)\right]'.$$

If B is an orthogonal complement of A, then $\Omega B = 0$, and

$$\begin{bmatrix} A' \\ B' \end{bmatrix} \Omega \begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

with $sep(D,0)=\lambda_{min}(D)$. Furthermore, following the proof of Theorem 1 of Lam et al. (2011), it holds that $d^{-2\delta}=O_p(\lambda_{min}(D))$, and $\| \hat{A}-A \| \leq \frac{4\| \hat{\Omega}-\Omega \|}{sep(D,0)}=O_p(d^{\delta}n^{-1/2})$.

Set $A=(\gamma_1,...,\gamma_r)$ and $\hat{A}=(\hat{\gamma}_1,...,\hat{\gamma}_r)$, then we have $\lambda_i=\gamma_i'\Omega\gamma_i$ and $\hat{\lambda}_i=\hat{\gamma}_i'\hat{\Omega}\hat{\gamma}_i$. Since $\| \hat{\gamma}_i-\gamma_i \| \leq \| \hat{A}-A \| = O_p(d^{\delta}n^{-1/2})$. For $i=1,...,r$, similar to the proof of Theorem 1 of Lam and Yao (2012), we decompose $\hat{\lambda}_i-\lambda_i=\hat{\gamma}_i'\hat{\Omega}\hat{\gamma}_i-\gamma_i'\Omega\gamma_i$, then we can obtain

$$| \hat{\lambda}_i-\lambda_i | = O_P(d^{-\delta}n^{-1/2}).$$

Also, similar to the proof of Theorem 1 of Lam and Yao (2012), we also have $\| \hat{B}-B \| = O_p(d^{\delta}n^{-1/2})$, where B is the orthogonal complement matrix. Set $B=(\gamma_{r+1},...,\gamma_d)$ and $\hat{B}=(\hat{\gamma}_{r+1},...,\hat{\gamma}_d)$. For $i=r+1,...,d$, we also decompose

$$\hat{\lambda}_i=\hat{\gamma}_i'\hat{\Omega}\hat{\gamma}_i=M_1+M_2+M_3+M_4,$$

where $M_1=(\hat{\gamma}_i-\gamma_i)'(\hat{\Omega}-\Omega)(\hat{\gamma}_i-\gamma_i), M_2=2(\hat{\gamma}_i-\gamma_i)'(\hat{\Omega}-\Omega)\gamma_i, M_3=(\hat{\gamma}_i-\gamma_i)'\Omega(\hat{\gamma}_i-\gamma_i), M_4=\gamma_i'(\hat{\Omega}-\Omega)\gamma_i$, then we can prove easily

$$| \hat{\lambda}_i | = O_P(n^{-1}).$$

Finally, similar to the proof of Corollary 1 of Lam and Yao (2012), on one hand, from the above proof, we have $d^{-2\delta}=O_p(\lambda_{min}(D))=\lambda_r$, and on the another hand,

$$\lambda_r \leq \lambda_1 = \| \Omega \| \leq \frac{1}{d^2}\sum_{l=1}^{l_0} \| \Sigma_Y(l) \|^2 = O(d^{-2\delta}).$$

Then, we have $\hat{\lambda}_i \asymp \lambda_i \asymp d^{-2\delta}$ for $i=1,...,r$, hence,

$$ER(i)\asymp 1 \quad \text{for} \quad i=1,...,r-1, \quad \text{and} \quad ER(r)=O_P(d^{2\delta}n^{-1}).$$

$\square$

Proof of Theorem 1. By the above proof of Lemma 1, for $i=1,...,r$, we have

$$\hat{\lambda}_i \asymp \lambda_i \asymp d^{-2\delta}, i=1,...,r.$$

Then,

$$\begin{aligned} CR(i) &= ER(i)\big(\sum_{k=i}^{m}\hat{\lambda}_k/\sum_{k=i+1}^{m}\hat{\lambda}_k\big) \\ &= ER(i)\big(1+\hat{\lambda}_i/\sum_{k=i+1}^{m}\hat{\lambda}_k\big), \quad i=1,...,r-1. \end{aligned}$$

When $md^{2\delta}n^{-1} \to \infty$ or $md^{2\delta}n^{-1} \to 0$, $1 + \hat{\lambda}_i / \sum_{k=i+1}^{m} \hat{\lambda}_k \asymp 1$. Also, $ER(i) \asymp 1$ for $i = 1, ..., r-1$. Hence,

$$CR(i) \asymp 1 \quad \text{for} \quad i = 1, ..., r-1.$$

Further, as $n = O(d^{2\delta+1})$, then

$$CR(r) = ER(r)\big(1 + \hat{\lambda}_r / \sum_{k=r+1}^{m} \hat{\lambda}_k\big) = O_P(d^{2\delta}n^{-1} + m^{-1}) = O_P(d^{2\delta}n^{-1}).$$

□

Proof of Corollary 1. As $n \to \infty$, $d \to \infty$ and $n = O(d^{2\delta+1})$, by the proof of Theorem 1, we have

$$CR(i) \asymp ER(i) \quad \text{for} \quad i = 1, ..., r-1,$$

and

$$\frac{CR(r)}{ER(r)} = O_P(1).$$

By the inequalities $c/(1+c) < ln(1+c) < c$, for $c > 0$, we have

$$ER(i) < GR(i) < ER(i)\big[1 + \big(\hat{\lambda}_i + \hat{\lambda}_{i+1}\big) / \sum_{k=i+2}^{m} \hat{\lambda}_k\big].$$

Thus, as $i = 1, ..., r-2$, we have

$$ER(i) < GR(i) < ER(i) \cdot O_P\big(1 + (1 + d^{2\delta})^{-1}\big),$$

as $i = r-1, r$, we have

$$ER(i) < GR(i) < ER(i) \cdot O_P(1 + d^{-2\delta}).$$

Hence,

$$GR(r) \asymp ER(r) \quad \text{for} \quad i = 1, ..., r.$$

This completes the proof of Corollary 1. □

Lemma 2. Under the Assumptions 1-8, and $\varrho_n = d^{\delta/2}n^{-1/2} \to 0$ and $n = O(d)$. As $n \to \infty$ and $d \to \infty$, then

(i) $\mid \hat{\lambda}_i - \lambda_i \mid = O_P(d^{-2\delta}\varrho_n)$ for $i = 1, ..., r$,

(ii) $|\hat{\lambda}_i| = O_P(d^{-\delta}n^{-1})$ for $i = r+1, ..., (l_0+1)r$,

(iii) $|\hat{\lambda}_i| = O_P(n^{-2})$ for $i = (l_0+1)r+1, ..., d$, and

(iv) $ER(i) \asymp 1$ for $i = 1, ..., r-1$, and $ER(r) = O_P(d^{\delta}n^{-1})$.

If in addition Assumption 9 holds, the rate in (ii) above can be further improved to

$$|\hat{\lambda}_i| = O_P(d^{-1/2-\delta}n^{-1}), \quad i = r+1, ..., (l_0+1)r,$$

and

$$ER(r) = O_P(d^{-1/2+\delta}n^{-1/2}).$$

Proof of Lemma 2. Under Assumptions 7 and 8, from the proof of Theorem 2 of Lam et al. (2011), $\| \hat{\Sigma}_{\xi}(l) \| = O_p(dn^{-1})$. Then, similar to the proof of Lemma 1, we have

$$\| \hat{\Omega} - \Omega \| = O_p(d^{-1-\delta}(d^{1-\delta/2}n^{-1/2} + dn^{-1})) = O_p(d^{-2\delta}\varrho_n).$$

For $i = 1, ..., r$, from the proof of Lemma 1, we have

$$|\hat{\lambda}_i - \lambda_i| = O_P(\| \hat{\Omega} - \Omega \|) = O_p(d^{-2\delta}\varrho_n),$$

which is Lemma 2 (i). Similar to the proof of Lemma 1, we also have $\| \hat{B} - B \| = O_p(\varrho_n)$, then, by the decomposition of $\hat{\lambda}_i = \hat{\gamma}_i'\hat{\Omega}\hat{\gamma}_i - \gamma_i'\Omega\gamma_i$, we can obtain

$$\hat{\lambda}_i = O_p(d^{-\delta}n^{-1}) \quad \text{for} \quad j = r+1, ..., (l_0+1)r,$$

which is Lemma 2 (ii).

To prove Lemma 2 (iii), we set $\boldsymbol{\Xi}_Y(l_0) = (\Sigma_Y(1), ..., \Sigma_Y(l_0))$ and $\hat{\boldsymbol{\Xi}}_Y(l_0) = (\hat{\Sigma}_Y(1), ..., \hat{\Sigma}_Y(l_0))$, then, we have

$$\Omega = \frac{1}{d^2}\boldsymbol{\Xi}_y(l_0)\boldsymbol{\Xi}_y'(l_0) \quad \text{and} \quad \hat{\Omega} = \frac{1}{d^2}\hat{\boldsymbol{\Xi}}_y(l_0)\hat{\boldsymbol{\Xi}}_y'(l_0).$$

Also, we set $\hat{\boldsymbol{\Xi}}_X(l_0) = (\hat{\Sigma}_X(1), ..., \hat{\Sigma}_X(l_0))$, $\hat{\boldsymbol{\Xi}}_{X\xi}(l_0) = (\hat{\Sigma}_{X\xi}(1), ..., \hat{\Sigma}_{X\xi}(l_0))$, $\hat{\boldsymbol{\Xi}}_{\xi X}(l_0) = (\hat{\Sigma}_{\xi X}(1), ..., \hat{\Sigma}_{\xi X}(l_0))$ and $\hat{\boldsymbol{\Xi}}_{\xi}(l_0) = (\hat{\Sigma}_{\xi}(1), ..., \hat{\Sigma}_{\xi}(l_0))$, then, we obtain

$$\hat{\boldsymbol{\Xi}}_Y(l_0) = A\big(\hat{\boldsymbol{\Xi}}_X(l_0)(I_{l_0} \otimes A') + \hat{\boldsymbol{\Xi}}_{X\xi}(l_0)\big) + \hat{\boldsymbol{\Xi}}_{\xi X}(l_0)(I_{l_0} \otimes A') + \hat{\boldsymbol{\Xi}}_{\xi}(l_0),$$

where $\otimes$ denotes and Kronecker product.

It is obvious that $rank\big(A\big(\hat{\boldsymbol{\Xi}}_X(l_0)(I_{l_0}\otimes A')+\hat{\boldsymbol{\Xi}}_{X\xi}(l_0)\big)+\hat{\boldsymbol{\Xi}}_{\xi X}(l_0)(I_{l_0}\otimes A')\big)\leq (l_0+1)r$, thus, $\sigma_i\big(A\big(\hat{\boldsymbol{\Xi}}_X(l_0)(I_{l_0}\otimes A')+\hat{\boldsymbol{\Xi}}_{X\xi}(l_0)\big)+\hat{\boldsymbol{\Xi}}_{\xi X}(l_0)(I_{l_0}\otimes A')\big)=0$ for $i=(l_0+1)r+1,...,d$, where $\sigma_i(M)$ denotes the i-th largest singular value of a matrix M.

Hence, for $i=(l_0+1)r+1,...,d$, we have

$$\lambda_i(\hat{\Omega})=\sigma_i^2(\frac{1}{d^2}\boldsymbol{\Xi}_Y(l_0))\leq\sigma_1^2(\frac{1}{d^2}\hat{\boldsymbol{\Xi}}_\xi(l_0))\leq\frac{1}{d^2}\sum_{i-1}^{l_0}\parallel\hat{\Sigma}_\xi(l)\parallel^2=O_p(n^{-2}).$$

According to the above conclusion, we have $\hat{\lambda}_i\asymp\lambda_i\asymp d^{-2\delta}$ for $i=1,...,r$, and $\hat{\lambda}_{r+1}=O_p(d^{-\delta}n^{-1})$, and Lemma 2 (iv) can be easily proved.

Finally, if Assumption 9 holds, we will prove $\mid\hat{\lambda}_i\mid=O_P(d^{-1/2-\delta}n^{-1}),\quad i=r+1,...,(l_0+1)r$. By Lemma 3 of Lam et al. (2011), with the same technique as in the proof of Theorem 1 in their paper, we have

$$\hat{B}=(B+AP)(I+P'P)^{-1/2}\quad\text{with}\quad\parallel P\parallel=O_p(\varrho_n).$$

According to the definition of $\hat{B}$ as in Lemma 1, then $\hat{\lambda}_{r+1}$ is the (1,1) element of the diagonal matrix $\hat{D}=\hat{B}'\hat{\Omega}\hat{B}$, where $\hat{\Omega}\hat{B}=\hat{B}\hat{D}$. Thus, we have $(I+P'P)^{1/2}B'\hat{\Omega}\hat{B}=\hat{D}$. Further, $\parallel P\parallel=O_p(\varrho_n)=o_p(1)$, hence, the rate of $\hat{\lambda}_{r+1}$ can be obtained using the (1,1) element of $B'\hat{\Omega}(B+AP)$. Similar to the proof of Theorem 2 of Lam and Yao (2012), the (1,1) element of $B'\hat{\Omega}(B+AP)$ has rate $O_P(d^{-1/2-\delta}n^{-1})$. Thus,

$$\mid\hat{\lambda}_i\mid=O_P(d^{-1/2-\delta}n^{-1}),\quad i=r+1,...,(l_0+1)r.$$

□

Proof of Theorem 2. Similar to Theorem 1, if Assumptions 1-8 hold, we have

$$CR(r)=ER(r)\big(1+\hat{\lambda}_r/\sum_{k=r+1}^m\hat{\lambda}_k\big)=O_P(d^\delta n^{-1}+m^{-1}).$$

If $n=O(d)$ and $d\geq n$, then we have

$$CR(r)=ER(r)\big(1+\hat{\lambda}_r/\sum_{k=r+1}^m\hat{\lambda}_k\big)=O_P(d^\delta n^{-1}).$$

Furthermore, if in addition Assumption 9 holds, then

$$CR(r)=ER(r)\big(1+\hat{\lambda}_r/\sum_{k=r+1}^m\hat{\lambda}_k\big)=O_P(d^{-1/2+\delta}n^{-1/2}+m^{-1}).$$

$\square$

Proof of Corollary 2. By Corollary 1, we have

$$ER(i) < GR(i) < ER(i)\big[1 + \big(\hat{\lambda}_i + \hat{\lambda}_{i+1}\big)/\sum_{k=i+2}^{m} \hat{\lambda}_k\big],$$

and then, as $i = 1, ..., r-2$, we have

$$ER(i) < GR(i) < ER(i) \cdot O_P\big(1 + (1 + d^{\delta})^{-1}\big),$$

as $i = r-1, r$, we have

$$ER(i) < GR(i) < ER(i) \cdot O_P(1 + d^{-\delta}).$$

Hence,

$$GR(r) \asymp ER(r) \quad \text{for} \quad i = 1, ..., r,$$

the proof of Corollary 2 is completed. $\square$

References

[1] Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica,* 81, 1203-1227.

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

[3] Amengual, D. and Watson, M. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business and Economic Statistics*, 25, 91-96.

[4] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70, 191-221.

[5] Chan, N. H., Lu, Y. and Yau, C. Y. (2017). Factor Modelling for High-Dimensional Time Series: Inference and Model Selection. *Journal of Time Series Analysis*, 38, 285-307.

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

[7] Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102, 603-617.

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

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

[10] Onatshi, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77, 1447-1479.

[11] Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economic and Statistics*, 92, 1004-1016.

[12] Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168, 244-258.

[13] Wu, J. (2016). Robust determination for the number of common factors in the approximate factor models. *Economics Letters*, 144, 102-106.

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

[15] Xia, Q., Xu, W. L. and Zhu, L. X. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statistica Sinica*, 25, 1025-1044.

Table 1: Relative frequency estimates for $P(\hat{r} = 3)$ with 200 reduplicate samples for Case 1

n	d	ER	RER	IC_1	ED	GR	CR
	0.2n	0.570(88|0)	0.460(108|0)	0.080(1|183)	0.350(0|130)	0.695(49|12)	0.755(24|25)
50	0.5n	0.510(98|0)	0.410(118|0)	0.350(0|130)	0.350(0|130)	0.660(53|15)	0.690(16|46)
	0.8n	0.530(93|1)	0.445(111|0)	0.250(0|150)	0.220(0|156)	0.720(32|24)	0.720(2|53)
	1.2n	0.615(77|0)	0.505(99|0)	0.210(0|158)	0.210(0|158)	0.785(13|30)	0.785(2|41)
	0.2n	0.825(35|0)	0.680(64|0)	0.815(0|37)	0.730(0|54)	0.955(6|3)	0.960(2|6)
100	0.5n	0.890(22|0)	0.800(40|0)	0.580(0|84)	0.440(0|112)	0.910(2|6)	0.900(0|10)
	0.8n	0.850(29|0)	0.785(43|0)	0.610(0|78)	0.355(0|129)	0.925(2|13)	0.920(0|16)
	1.2n	0.835(33|0)	0.740(50|0)	0.255(0|149)	0.085(0|183)	0.895(3|18)	0.875(0|25)
	0.2n	0.995(1|0)	0.965(7|0)	0.960(0|8)	0.800(0|40)	1.000(0|0)	1.000(0|0)
200	0.5n	0.975(5|0)	0.960(8|0)	0.685(0|63)	0.395(0|121)	0.990(0|2)	0.990(0|2)
	0.8n	0.985(3|0)	0.945(11|0)	0.465(0|110)	0.115(0|177)	0.990(0|2)	0.985(0|3)
	1.2n	0.990(2|0)	0.950(10|0)	0.260(0|148)	0.070(0|186)	0.990(0|2)	0.990(0|2)
	0.2n	1.000(0|0)	1.000(0|0)	0.985(0|3)	0.805(0|39)	1.000(0|0)	1.000(0|0)
500	0.5n	1.000(0|0)	1.000(0|0)	0.835(0|33)	0.140(0|172)	1.000(0|0)	1.000(0|0)
	0.8n	1.000(0|0)	1.000(0|0)	0.655(0|69)	0.030(0|194)	1.000(0|0)	1.000(0|0)
	1.2n	1.000(0|0)	1.000(0|0)	0.380(0|124)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	0.2n	1.000(0|0)	1.000(0|0)	1.000(0|0)	0.755(0|49)	1.000(0|0)	1.000(0|0)
1000	0.5n	1.000(0|0)	1.000(0|0)	0.980(0|4)	0.025(0|195)	1.000(0|0)	1.000(0|0)
	0.8n	1.000(0|0)	1.000(0|0)	0.850(0|26)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	1.2n	1.000(0|0)	1.000(0|0)	0.560(0|88)	0.000(0|200)	1.000(0|0)	1.000(0|0)

Table 2: Relative frequency estimates for $P(\hat{r} = 3)$ with 200 reduplicate samples for Case 2

n	d	ER	RER	IC_1	ED	GR	CR
	0.2n	0.290(140|2)	0.200(160|0)	0.570(8|78)	0.250(3|147)	0.430(98|16)	0.525(80|15)
50	0.5n	0.325(134|1)	0.210(158|0)	0.725(5|50)	0.595(8|73)	0.530(90|4)	0.610(66|12)
	0.8n	0.310(138|0)	0.165(167|0)	0.670(2|64)	0.480(5|99)	0.530(92|2)	0.635(65|8)
	1.2n	0.305(139|0)	0.210(158|0)	0.685(2|61)	0.465(3|114)	0.55081|9)	0.650(55|15)
	0.2n	0.520(96|0)	0.320(139|0)	0.865(0|27)	0.770(0|46)	0.835(33|0)	0.900(18|2)
100	0.5n	0.605(79|0)	0.300(120|0)	0.890(0|22)	0.625(0|75)	0.870(26|0)	0.920(16|0)
	0.8n	0.525(94|1)	0.195(151|0)	0.810(0|38)	0.455(0|109)	0.860(26|2)	0.910(13|5)
	1.2n	0.485(103|0)	0.200(160|0)	0.760(0|48)	0.285(0|153)	0.865(27|0)	0.920(13|3)
	0.2n	0.810(38|0)	0.450(110|0)	0.980(0|4)	0.800(0|40)	0.985(3|0)	0.990(1|1)
200	0.5n	0.795(41|0)	0.250(150|0)	0.860(0|28)	0.440(0|112)	0.920(6|0)	0.980(3|1)
	0.8n	0.855(29|0)	0.250(150|0)	0.860(0|28)	0.265(0|147)	0.980(4|0)	0.990(1|1)
	1.2n	0.750(50|0)	0.180(164|0)	0.680(0|64)	0.110(0|178)	0.970(6|0)	0.990(1|1)
	0.2n	1.000(0|0)	0.640(72|0)	0.995(0|1)	0.870(0|26)	1.000(0|0)	1.000(0|0)
500	0.5n	1.000(0|0)	0.370(126|0)	0.965(0|7)	0.195(0|161)	1.000(0|0)	1.000(0|0)
	0.8n	1.000(0|0)	0.245(151|0)	0.940(0|12)	0.030(0|194)	1.000(0|0)	1.000(0|0)
	1.2n	1.000(0|0)	0.105(179|0)	0.795(0|41)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	0.2n	1.000(0|0)	0.860(28|0)	1.000(0|0)	0.835(0|33)	1.000(0|0)	1.000(0|0)
1000	0.5n	1.000(0|0)	0.525(95|0)	0.995(0|1)	0.025(0|195)	1.000(0|0)	1.000(0|0)
	0.8n	1.000(0|0)	0.250(149|0)	0.935(0|3)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	1.2n	1.000(0|0)	0.090(182|0)	0.870(0|26)	0.000(0|200)	1.000(0|0)	1.000(0|0)

Table 3: Relative frequency estimates for $P(\hat{r} = 3)$ with 200 reduplicate samples for Case 3

n	d	ER	RER	IC_1	ED	GR	CR
	0.2n	0.190(162|0)	0.145(171|0)	0.490(7|95)	0.480(12|92)	0.475(103|2)	0.585(69|12)
50	0.5n	0.120(176|1)	0.085(183|0)	0.665(6|61)	0.585(3|80)	0.400(120|0)	0.615(71|6)
	0.8n	0.095(181|0)	0.050(190|0)	0.670(2|64)	0.525(7|88)	0.370(124|2)	0.535(88|5)
	1.2n	0.070(186|0)	0.020(196|0)	0.585(1|82)	0.340(2|130)	0.335(132|1)	0.555(82|7)
	0.2n	0.270(146|0)	0.130(174|0)	0.945(0|11)	0.795(2|39)	0.805(39|0)	0.900(20|0)
100	0.5n	0.180(164|0)	0.060(188|0)	0.850(4|26)	0.630(0|74)	0.760(48|0)	0.885(22|1)
	0.8n	0.105(179|1)	0.015(197|0)	0.690(0|62)	0.405(0|119)	0.635(73|0)	0.835(32|1)
	1.2n	0.090(182|0)	0.010(198|0)	0.660(0|88)	0.240(0|152)	0.665(87|0)	0.780(42|2)
	0.2n	0.340(132|0)	0.095(181|0)	0.970(0|6)	0.855(0|29)	0.965(7|0)	0.995(0|1)
200	0.5n	0.280(144|0)	0.005(199|1)	0.860(0|28)	0.465(0|107)	0.905(19|0)	0.975(4|1)
	0.8n	0.130(174|0)	0.005(199|0)	0.785(0|43)	0.220(0|156)	0.860(28|0)	0.975(5|0)
	1.2n	0.100(180|0)	0.000(200|0)	0.735(0|64)	0.100(0|180)	0.865(27|0)	0.980(4|0)
	0.2n	0.595(81|0)	0.015(197|0)	1.000(0|0)	0.875(0|35)	1.000(0|0)	1.000(0|0)
500	0.5n	0.430(114|0)	0.000(200|0)	0.990(0|2)	0.220(0|156)	1.000(0|0)	1.000(0|0)
	0.8n	0.215(157|0)	0.245(151|0)	0.865(0|25)	0.030(0|194)	1.000(0|0)	1.000(0|0)
	1.2n	0.125(175|0)	0.000(200|0)	0.800(0|40)	0.005(0|199)	0.995(1|0)	1.000(0|0)
	0.2n	0.865(27|0)	0.000(200|0)	1.000(0|2)	0.775(0|45)	1.000(0|0)	1.000(0|0)
1000	0.5n	0.705(59|0)	0.000(200|0)	1.000(0|0)	0.020(0|196)	1.000(0|0)	1.000(0|0)
	0.8n	0.415(117|0)	0.000(200|0)	0.0.985(0|3)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	1.2n	0.240(152|0)	0.000(200|0)	0.870(0|26)	0.000(0|200)	1.000(0|0)	1.000(0|0)

Table 4: Relative frequency estimates for $P(\hat{r} = 3)$ with 200 reduplicate samples for Case 4

n	d	ER	RER	IC_1	ED	GR	CR
50	0.2n	0.180(163|1)	0.100(179|1)	0.525(34|11)	0.450(13|9)	0.475(103|2)	0.495(88|13)
	0.5n	0.095(181|10)	0.050(190|0)	0.645(8|63)	0.565(3|84)	0.400(120|0)	0.610(67|11)
	0.8n	0.075(185|0)	0.015(197|0)	0.730(2|52)	0.570(2|84)	0.390(122|0)	0.655(64|5)
	1.2n	0.045(191|0)	0.005(199|0)	0.655(3|66)	0.420(1|115)	0.335(133|0)	0.580(68|16)
100	0.2n	0.060(188|0)	0.100(198|0)	0.850(0|30)	0.765(0|47)	0.785(43|0)	0.895(16|5)
	0.5n	0.075(185|0)	0.000(200|0)	0.850(0|30)	0.615(0|77)	0.750(50|0)	0.890(17|5)
	0.8n	0.055(189|1)	0.010(198|0)	0.785(0|43)	0.420(0|116)	0.730(54|0)	0.900(12|8)
	1.2n	0.020(196|0)	0.000(200|0)	0.720(0|56)	0.270(0|146)	0.655(68|1)	0.885(9|14)
200	0.2n	0.095(181|0)	0.000(200|0)	0.975(0|5)	0.805(0|39)	0.975(5|0)	0.990(1|1)
	0.5n	0.025(195|0)	0.000(200|0)	0.905(0|19)	0.450(0|110)	0.975(5|0)	0.990(1|1)
	0.8n	0.010(198|0)	0.000(200|0)	0.825(0|35)	0.255(0|149)	0.970(6|0)	0.995(0|1)
	1.2n	0.050(199|0)	0.000(200|0)	0.715(0|57)	0.085(0|183)	0.940(12|0)	0.980(1|3)
500	0.2n	0.150(169|0)	0.000(200|0)	0.995(0|1)	0.870(0|26)	1.000(0|0)	1.000(0|0)
	0.5n	0.030(194|0)	0.000(200|0)	0.995(0|1)	0.200(0|159)	1.000(0|0)	1.000(0|0)
	0.8n	0.005(199|0)	0.000(200|0)	0.900(0|20)	0.030(0|194)	1.000(0|0)	1.000(0|0)
	1.2n	0.000(200|0)	0.000(200|0)	0.800(0|40)	0.005(0|199)	1.000(0|0)	1.000(0|0)
1000	0.2n	0.270(146|0)	0.000(200|0)	1.000(0|0)	0.775(0|45)	1.000(0|0)	1.000(0|0)
	0.5n	0.005(199|0)	0.000(200|0)	1.000(0|0)	0.035(0|193)	1.000(0|0)	1.000(0|0)
	0.8n	0.005(199|0)	0.000(200|0)	0.970(0|6)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	1.2n	0.000(200|0)	0.000(200|0)	0.905(0|19)	0.000(0|200)	1.000(0|0)	1.000(0|0)

Table 5: Relative frequency estimates for $P(\hat{r}=3)$ with 200 reduplicate samples for Case 5

n	d	ER	RER	IC_1	ED	GR	CR
	0.2n	0.135(173|1)	0.090(182|0)	0.550(8|82)	0.555(9|80)	0.380(103|2)	0.575(73|12)
50	0.5n	0.075(185|0)	0.035(193|0)	0.645(7|64)	0.500(2|98)	0.370(120|0)	0.660(67|11)
	0.8n	0.065(187|0)	0.025(195|0)	0.585(3|80)	0.465(1|106)	0.360(128|0)	0.610(64|5)
	1.2n	0.025(195|0)	0.015(197|0)	0.465(3|104)	0.320(1|135)	0.285(141|2)	0.590(58|24)
	0.2n	0.110(178|0)	0.045(191|0)	0.875(0|25)	0.770(0|46)	0.700(59|1)	0.905(11|8)
100	0.5n	0.020(196|0)	0.005(199|0)	0.675(0|65)	0.450(0|110)	0.650(70|0)	0.865(11|16)
	0.8n	0.025(195|0)	0.000(200|0)	0.615(1|76)	0.320(1|135)	0.665(67|0)	0.930(5|9)
	1.2n	0.010(198|0)	0.000(200|0)	0.435(0|113)	0.135(0|173)	0.585(81|2)	0.900(3|17)
	0.2n	0.075(185|0)	0.005(199|0)	0.950(0|10)	0.785(0|43)	0.970(6|0)	1.000(0|0)
200	0.5n	0.025(195|0)	0.000(200|0)	0.815(0|37)	0.390(0|122)	0.965(7|0)	0.995(0|1)
	0.8n	0.010(198|0)	0.000(200|0)	0.665(0|67)	0.170(0|166)	0.905(19|0)	0.990(0|2)
	1.2n	0.000(200|0)	0.000(200|0)	0.510(0|98)	0.025(0|195)	0.905(19|0)	0.975(0|5)
	0.2n	0.075(185|0)	0.000(200|0)	1.000(0|0)	0.795(0|41)	1.000(0|0)	1.000(0|0)
500	0.5n	0.020(196|0)	0.000(200|0)	0.910(0|18)	0.165(0|167)	1.000(0|0)	1.000(0|0)
	0.8n	0.000(200|0)	0.000(200|0)	0.800(0|40)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	1.2n	0.000(200|0)	0.000(200|0)	0.645(0|71)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	0.2n	0.290(142|0)	0.000(200|0)	1.000(0|0)	0.765(0|47)	1.000(0|0)	1.000(0|0)
1000	0.5n	0.015(197|0)	0.000(200|0)	0.985(0|3)	0.020(0|196)	1.000(0|0)	1.000(0|0)
	0.8n	0.005(199|0)	0.000(200|0)	0.920(0|16)	0.000(0|200)	1.000(0|0)	1.000(0|0)
	1.2n	0.000(200|0)	0.000(200|0)	0.840(0|32)	0.000(0|200)	1.000(0|0)	1.000(0|0)

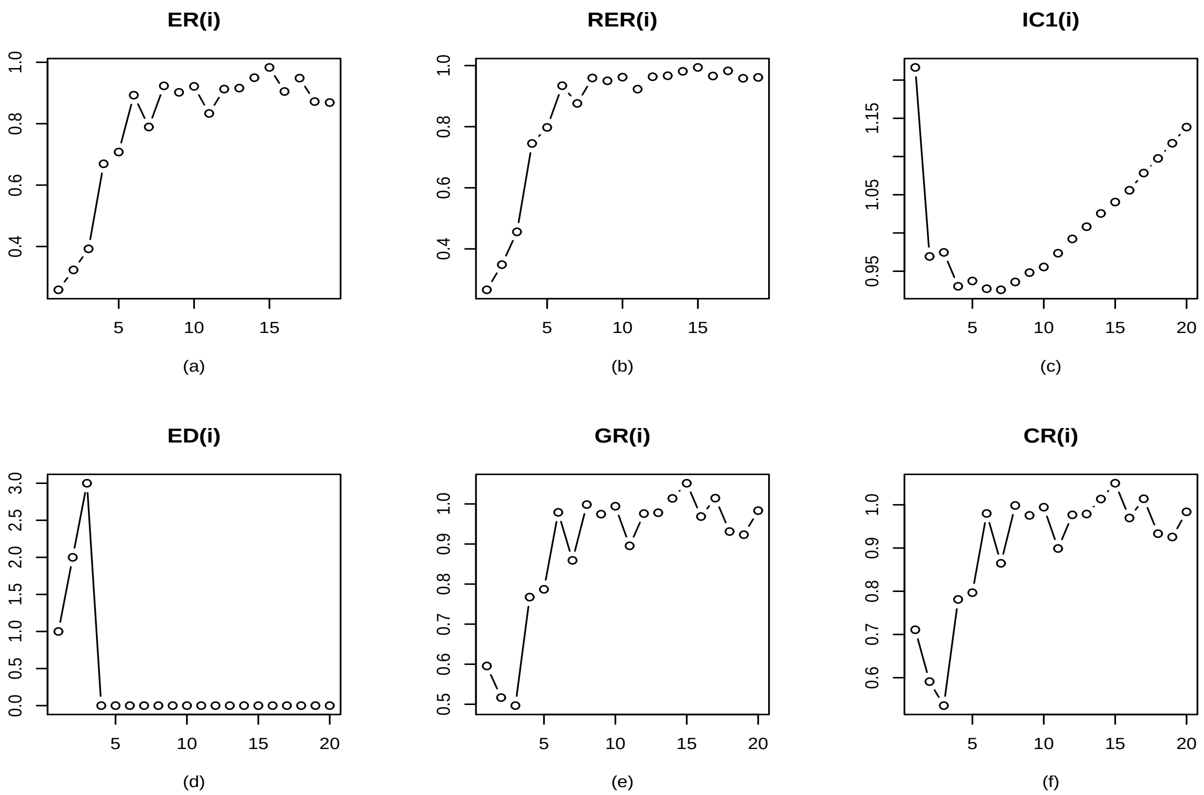

Figure 1: Plots of ER (a), RER (b), IC1 (c), ED (d), GR (e) and CR (f) for the real data set.