Statistica Sinica **25** (2015), 1025-1044
doi:http://dx.doi.org/10.5705/ss.2013.252

CONSISTENTLY DETERMINING THE NUMBER OF FACTORS IN MULTIVARIATE VOLATILITY MODELLING

Qiang Xia[1], Wangli Xu[1] and Lixing Zhu[2]

[1]*Renmin University of China and* [2]*Hong Kong Baptist University*

Abstract: Consistently determining the number of factors plays an important role in factor modelling for volatility of multivariate time series. In this paper, the modelling is extended to handle the nonstationary time series scenario with conditional heteroscedasticity. Then a ridge-type ratio estimate and a BIC-type estimate are proposed and proved to be consistent. Their finite sample performance is examined through simulations and the analysis of two data sets. An observation from the numerical studies is, that unlike the cases with stationary and homoscedastic sequences in the literature, the dimensionality blessing no longer holds for the ratio-based estimates, but still does for the BIC-type estimate.

Key words and phrases: BIC-type criterion, dimension reduction, eigenanalysis, factor modelling, multivariate volatility, nonstationarity, ratio estimate.

1. Introduction

Over the last two decades, the modelling of time-varying volatility of financial returns has been a vigorous area of research in the financial econometrics literatures. Many statistical models, mostly designed for univariate data, have been proposed. The analysis of multivariate time series data, often through the modelling of multivariate processes with conditional covariance matrices, has motivated the attempts to extend univariate volatility models to multivariate cases. For example, Engle and Kroner (1995) proposed the general multivariate GARCH(p, q) model with the BEKK representation. Although this model is general, it cannot tackle the overparametrization problem, which is similar to those for multivariate ARMA models. These models are seldomly used directly in practice. To overcome the difficulties caused by overparametrization for multivariate GARCH(p, q) models, there are such proposals as those of Engle (2002), Fan, Wang, and Yao (2008), and Pan et al. (2011), among others.

An effective way to circumvent the aforementioned problem is to adopt a factor model that provides a low-dimensional parsimonious representation for high-dimensional dynamics. Early attempts in this direction include Anderson (1963), Priestley, Rao, and Tong (1974), and Peña and Box (1987). Recently, more efforts focus on the inference when the dimension of time series d is as large as or greater than the sample size n; see, for example, Bai (2003), Forni et al. (2000, 2004, 2005), Lam, Yao, and Bathia (2011) and Lam and Yao (2012). Relevant efforts look to building a factor model for nonstationary multivariate time series, for example, Peña and Poncela (2006) and Pan and Yao (2008).

Consistently determining the number of factors plays an important role in establishing a factor model. A ratio estimate was suggested by Lam and Yao (2012), motivated from a similar idea suggested by Wang (2012) under an *iid* setup; in their numerical studies the ratio works very well. They also found that dimensionality blesses the estimation accuracy of the ratio estimate, but the selection consistency remains unknown even when the components of X are white noises. Under normality and other conditions, the ratio estimate in Wang (2012) can be consistent when p is much larger than the sample size n. One condition plays a critical role in establishing the consistency. And, when not satisfied, the method there may not be applicable for us. If a consistence of Lam and Yao (2012) were true in this condition, consistency would be easily derived. Based on numerical study, this seems not easy to derive. To define a consistent estimate, we suggest a ridge-type ratio estimation method that is a modification of the ratio estimate, but does not require, up to certain order moments, any distributional assumptions. Further, another eigenvalues-based criterion of BIC-type is suggested to consistently determine the number of factors in the setup under study.

To assess the finite-sample performance of the two estimates, simulations were carried out. The ratio-based estimates worked well in many cases. On the other hand, the numerical studies suggests that, unlike the finding of dimensionality blessing by Lam and Yao (2012), the ratio-based estimates do not have this property for nonstationary and heteroscedastic time series, while the BIC-type estimate does.

The rest of the paper is arranged as follows. The methodology development and asymptotic properties of the proposed estimates are described in Section 2. Simulated and data examples are presented in Section 3. Some conclusions are included in Section 4. The technical arguments are contained in the online supplement.

Throughout the paper, τ stands for the transpose. For any matrix M, $\| M \|_{max}$ and $\| M \|_{min}$ denote the positive square roots of the maximum and minimum eigenvalue of MM^τ and $M^\tau M$, respectively. When M is a vector, $\| M \|_{max} = \| M \|_{min}$ is equivalent to the usual L_2-norm. When $a = O_p(b)$ and $b = O_p(a)$, we write $a \asymp b$.

2. Model and Estimation

2.1. A brief review of the volatility factor model

For $t = 1, \ldots, n$, let Y_t be an d-dimensional vector of observed time series with $E(Y_t|\mathcal{F}_{t-1}) = \mu_t$, where $\mathcal{F}_t = \sigma(Y_t, Y_{t-1}, \ldots)$ and μ_t is related to t, and with the conditional covariance matrices $\Sigma_t = Var(Y_t|\mathcal{F}_{t-1})$ for $t = 1, \ldots, n$. To overcome difficulties due to overparametrization (Fan, Wang, and Yao (2008)), a common factor model is used, as suggested by Pan et al. (2011),

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

where $X_t = (x_{1t}, \ldots, x_{rt})^\tau$ is the r-dimensional vector of nonobserved (nonstationary) common factor, $r < d$ is unknown; A is an $d \times r$ factor loading matrix; $\{\xi_t\}$ is a sequence of *i.i.d.* innovations with mean $\mathbf{0}$ and Σ_ξ, and $\{\xi_t\}$ is independent of X_t and $\mathcal{F}_{t-1}$. The volatility dynamics structure of Y_t here comes through the lower dimensional volatility dynamics of X_t, and the static variate of ξ_t,

$$\Sigma_t = A\Sigma_{X_t}A^\tau + \Sigma_\xi. \tag{2.2}$$

The model (2.1) is not identifiable: for any $r \times r$ nonsingular H, Y_t can be expressed by a new factor as

$$Y_t = AHH^{-1}X_t + \xi_t. \tag{2.3}$$

However the linear space $\mathcal{M}(A)$, the factor loading space, spanned by the columns of A is uniquely defined by (2.1). The factor process X_t is also uniquely defined accordingly so, for example, we may choose a half orthogonal matrix A, with $A^\tau A = I_r$, where I_r is the $r \times r$ identity matrix. Thus, one can rotate an estimated factor loading matrix whenever appropriate, and this can facilitate the estimation of A.

2.2. Methodology development

To include nonstationary cases in our asymptotic theory, we require some regularity conditions. Let $\Sigma_Y(m,t) = Cov(Y_{t+m}, Y_t), \Sigma_X(m,t) = Cov(X_{t+m}, X_t)$, and $\Sigma_{X\xi}(m,t) = Cov(X_{t+m}, \xi_t)$, for all $m \geq 1$. Under stationarity, we write $\Sigma_Y(m)$, $\Sigma_X(m)$, and $\Sigma_{X\xi}(m)$, respectively.

When our Condition 4 holds, $\Sigma_Y(m) = A\Sigma_X(m)A^\tau + A\Sigma_{X\xi}(m)$. In light of Lam and Yao (2012), we take

$$\Omega = \frac{1}{d^2}\sum_{m=1}^{m_0} \Sigma_Y(m)\Sigma_Y^\tau(m),$$

where m_0 is a prescribed positive integer. This helps us construct consistent estimations of the number of factors later. As Ω is nonnegative definite, the eigenvalues are real and nonnegative. Further, when $\Omega B = 0$, $\Sigma_Y^\tau(m)B = 0$ for all $m \geq 1$. The number of factors r is then the number of non-zero eigenvalues λ_i of Ω, the loading matrix A is comprised of the corresponding r orthonormal eigenvectors α_i. Hence, to estimate the number of factors r and the loading matrix A, performing an eigenanalysis for the sample counterparts of Ω is needed:

$$\hat{\Omega} = \sum_{m=1}^{m_0} S_m S_m^\tau, \tag{2.4}$$

where $S_m = (1/[(n-m)\cdot d])\sum_{t=1}^{n-m}(Y_{t+m}-\bar{Y})(Y_t-\bar{Y})^\tau$ and $\bar{Y} = (1/n)\sum_{t=1}^{n} Y_t$.

Remark 1. Since $\Sigma_Y(0) = A\Sigma_X(0)A^\tau + \Sigma_\xi(0)$ and $\Sigma_Y(0)B \neq 0$, $m = 0$ is excluded from the sum in Ω, and B is an orthogonal complement matrix of A.

We need some regularity conditions similar to those in Lam, Yao, and Bathia (2011), Lam and Yao (2012), and Pan and Yao (2008).

Condition 1. $\Sigma_X(m,t)$ and $\Sigma_{X\xi}(m,t)$ are free of t, say $\Sigma_X(m)$ and $\Sigma_{X\xi}(m)$.

Condition 2. $\| \Sigma_X(m) \|_{max} \asymp \| \Sigma_X(m) \|_{min} \asymp d$, for $m = 0, 1, \ldots, m_0$, where m_0 is a positive integer.

Condition 3. $\Sigma_{X\xi}(m)$ remains bounded in all elements.

Condition 4. The covariance matrix $Cov(\xi_t, X_s) = 0$ for all $t \geq s$.

Condition 5. The $Y_{t,i}$'s are nonstationary φ-mixing with $\varphi(l) = O(l^{-2-v})$ for some positive constant v, where

$$\varphi(l) = \sup_{k\geq 1} \sup_{U\in\mathcal{F}_{-\infty}^k, V\in\mathcal{F}_{l+k}^\infty, P(V)>0} | P(V|U) - P(V) | \tag{2.5}$$

and $\mathcal{F}_i^j = \sigma(Y_i, \ldots, Y_j)$. Further, $\max_{t,i} E|Y_{t,i}|^4 < \infty$.

Condition 6. $\sqrt{n}(E(dS_m) - \Sigma_Y(m)) = O(1)$ elementwisely in the sense that $O(1)$ is uniformly for all the elements.

Remark 2. Condition 2 is similar to condition C5 of Lam and Yao (2012) with $\delta = 0$. Condition 3 and 4 are in common use. Condition 6 is similar to that in Theorem 2 of Pan and Yao (2008) in which d is fixed and $\sqrt{n}(E(d\cdot S_m) - \Sigma_Y(m)) = O(1)$ when our notations are adopted.

Theorem 1. *Let Conditions* $1 \sim 6$ *hold, and assume that the* r *largest eigenvalues* $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = 0 = \cdots = \lambda_d = 0$ *of* Ω *are distinct. Then*

(i) $\| \hat{\alpha}_i - \alpha_i \| \leq \| \hat{A} - A \| = O_P(n^{-1/2})$,

(ii) $| \hat{\lambda}_i - \lambda_i | = O_P(n^{-1/2})$ *for* $i = 1, \ldots, r$,

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

Here $\hat{A} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_r)$, *where the* $\hat{\alpha}_i$ *are the eigenvectors of* $\hat{\Omega}$ *corresponding to the* i *largest eigenvalues. Now estimating* B *is equivalent to estimating* A, *where* $B = (\beta_1, \ldots, \beta_{d-r})$ *is a* $d \times (d-r)$ *matrix with* $B^\tau A = \mathbf{0}$, $B^\tau B = I_{d-r}$ *and* $AA^\tau + BB^\tau = I_d$.

Due to the random fluctuation in a sample of finite size, the estimates of the zero-eigenvalues of Ω are unlikely to be 0. Hence, we cannot directly determine r from the first nonzero-eigenvalues. Some methods for determining r can be found in the literature. In a different setting, Lam and Yao (2012) suggested a ratio estimate of r, using a similar idea as that in Wang (2012). An estimate $\hat{r}$ is

$$\hat{r} = \arg \min_{k \in \{1,2,\ldots,R\}} \frac{\hat{\lambda}_{k+1}}{\hat{\lambda}_k}, \tag{2.6}$$

where $\hat{\lambda}_d \leq \cdots \leq \hat{\lambda}_1$ are the eigenvalues of $\hat{\Omega}$, and $r < R < d$ is a constant. In numerical studies this criterion has been proved to work well, mainly because of the fast convergence rate of $\hat{\lambda}_k$ to zero for $k > r$. They chose $R = d/2$ to avoid poor ratio estimates $\hat{\lambda}_{k+1}/\hat{\lambda}_k$ for large k. However, the estimate may not be stable. Lam and Yao (2012) conjectured that $\hat{\lambda}_{k+1}/\hat{\lambda}_k$ would converge to 1 as $k \to \infty$. However, this is a challenging task and, from Figure 3, 6, and 10 in Section 3, it seems not easy to verify. Also, when d is large, $R = d/2$ may be too large when d is large.

In this paper, we propose two estimates without the strong conditions assumed in the literature to produce consistency.

Ridge-type ratio estimate (*RRE*). This estimate is a modification achieved by adding a positive value c to the eigenvalues λ_i,

$$\hat{r} = \arg \min_{k \in \{1,2,\ldots,d\}} \frac{\hat{\lambda}_{k+1} + c}{\hat{\lambda}_k + c}. \tag{2.7}$$

When $k \geq r$ and c is chosen to be larger than $\hat{\lambda}_{k+1}$, the minimum over $k = 1, \ldots, d$ is equivalent to the minimum over $k = 1, \ldots, R$. From Theorem 1 and our choice of c, it is true in a probability sense. In our simulations, we computed the latter minimum for convenience. This is an idea similar to adding a ridge. The choice of c requires care, see Theorem 2 about this.

BIC-type estimate. This method is applicable because in the sufficient dimension reduction, the corresponding eigenvectors that are associated with the nonzero eigenvalues of a target matrix are used as the base vectors of the space we want to determine, see Cook (1998), and relevant references are in Zhu, Miao, and Peng (2006), Zhu and Zhu (2009), and Feng et al. (2013). The method depends on the eigenvalues $\hat{\lambda}_i$ as in

$$\hat{r} = \arg \max_{k \in \{1,2,\ldots,d\}} \left(\frac{n \sum_{m=1}^{k} (\log(\hat{\lambda}_m + 1) - \hat{\lambda}_m)}{2 \sum_{m=1}^{d} (\log(\hat{\lambda}_m + 1) - \hat{\lambda}_m)} - 2 \log n \times \frac{k(k+1)/2}{d} \right), \tag{2.8}$$

where $\hat{\lambda}_d \leq \cdots \leq \hat{\lambda}_1$, and $k(k+1)/2$ is the number of free parameters. Here, the maximum can be taken over all eigenvalues. The penalty used, $2 \log n$, is the one used in the classical BIC; the criterion proposed by Zhu, Miao, and Peng (2006) needs a delicately selected penalty. It seems that for factor modelling, the BIC-type estimate is less sensitive to penalty selection than the one used for generic dimension reduction by Zhu, Miao, and Peng (2006) studied.

Theorem 2. *Under the conditions of Theorem* 1*, the estimates* $\hat{r}$ *of* (2.7) *with* $c = \log n/(10n)$*, and* $\hat{r}$ *of* (2.8) *with* $d/(n \log n) \to 0$ *satisfy* $P(\hat{r} = r) \to 1$ *as* $n \to \infty$.

From the proof, for a relatively wide range of c, consistency holds. Based on the limited simulation experiments we have conducted, $c = \log n/(10n)$ is a recommended value for practitioners.

3. Numerical Studies

We carried out three simulation experiments to examine the performance of the estimates of r and A. The estimates of A and r are not sensitive to the choice of m_0 (Lam, Yao, and Bathia (2011); Lam and Yao (2012)), so we set $m_0 = 1$ in all the simulations. These results were then used to model multivariate volatilities for a data example.

3.1. Simulation experiments

Example 1. Consider a simple simulated model with one factor: $r = 1$, the $d \times r$ matrix is a vector $A = (1, \ldots, 1)^{\tau}$, and the ξ_t are $N(\mathbf{0}, I_d)$ independent variables. The factor was taken as

$$x_t - \frac{t}{n} = 0.5\Big(x_{t-1} - \frac{t}{n}\Big) + \varepsilon_t, \quad \varepsilon_t = \sigma_t e_t, \quad \sigma_t^2 = 0.3 + 0.7\varepsilon_{t-1}^2, \tag{3.1}$$

where $\{e_t\}$ is a sequence of independent $N(0, 1)$ random variables.

The factor is non-stationary with nonconstant means over t and conditional heteroscedasticity. In the simulation, the sample size was $n = 50, 100, 200, 500$, and 1,000, and the dimension was $d = n/2$. 200 samples werw drawn in every setting. Figures 1, 2, 3, 4, and 5, respectively, depict the boxplots for the errors of $\hat{\lambda}_i - \lambda_i, i = 1, \ldots, 6$, $\| \hat{A} - A \|$, the ratios $\hat{\lambda}_{i+1}/\hat{\lambda}_i$, the ridge-type ratios $(\hat{\lambda}_{i+1} + \log(n)/10n)/(\hat{\lambda}_i + \log(n)/10n)$, and the BIC-type with $C_n = \log(n)$. In Figure 1, the estimation errors of $\hat{\lambda}_i - \lambda_i, i = 1, \ldots, 6$ seem to converge to 0 as both n and d get larger, suggesting consistency. Figure 2 shows that the estimation errors for the factor loading coefficients go to 0 as both n and d increase. Although the ratio estimate, the ridge-type ratio estimate and the BIC-type estimate for the number of factors work well, the results in Table 1 are that the ridge-type ratio estimate is the best with the highest percentage of correctly determined

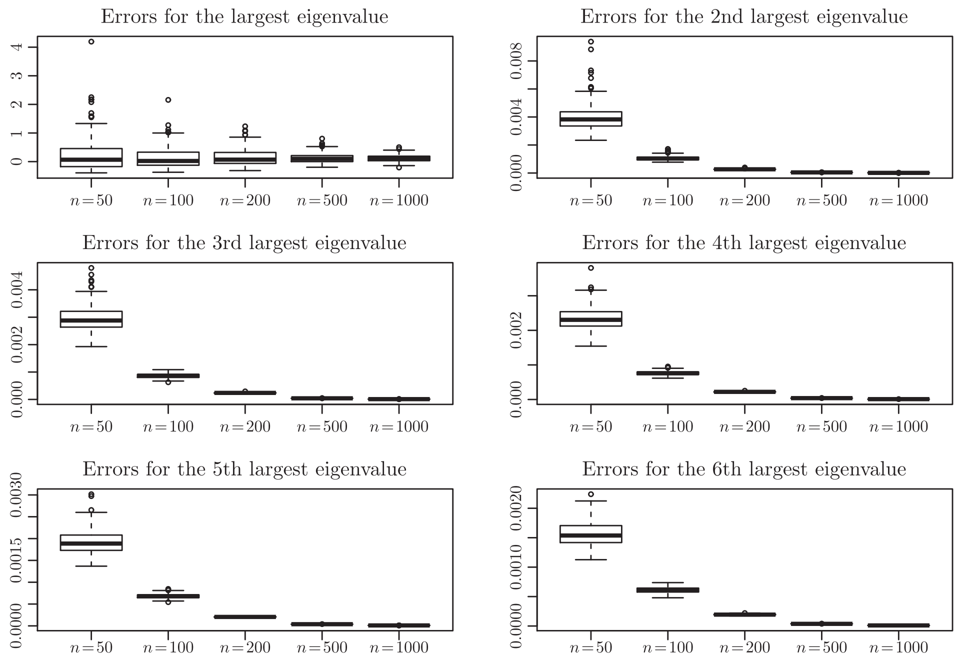

Figure 1. Boxplots for errors in estimating the first six eigenvalues of Ω with $r = 1$ and $d = n/2$ for model (3.1).

Table 1. Example 1: Relative frequency estimates for $P(\hat{r} = 1)$ with 200 replicated samples in the simulation.

n	d	RE	RRE	BIC
50	25	0.775	0.995	0.64
100	50	0.925	1	0.76
200	100	0.975	1	0.90
500	250	0.99	1	1
1,000	500	0.99	1	1

number of factors in the small n cases (n=50, 100 and 200), and the BIC-type estimates are the worst. However, with n=500 or 1,000, the BIC-type estimate are as good as the ridge-type ratio estimate, and both are slightly better than the ratio estimate. This might be due to the fact that the ratios for large i are instable as they are 0/0 at population level. Figure 3 suggests this. This seems to show that the conjecture by Lam and Yao (2012) may not be true. But for the estimate of r, even when the sample size is as small as $n = 50$, Figures 3, 4, and 5 suggest that $\hat{r}$ may converge to 1, this behavior is the same as that observed by Lam and Yao (2012) when they considered one factor for stationary time series.

Example 2. Here $r = 3$, and all the elements of $A_{d\times r}$ were generated in-

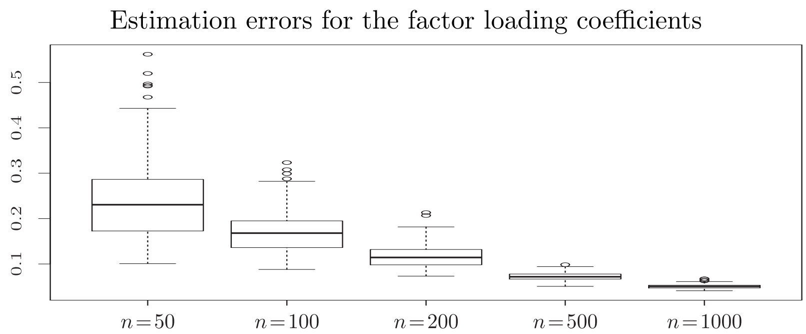

Figure 2. Boxplots for ($\| \hat{A} - A \|$) for model (3.1).

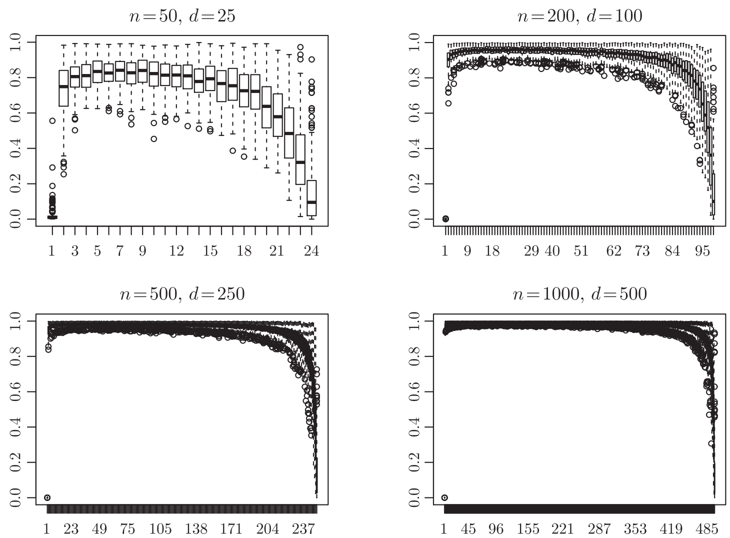

Figure 3. Boxplots for the ratios $\hat{\lambda}_{i+1}/\hat{\lambda}_i, i = 1, \ldots, d$, for model (3.1).

dependently from the uniform distribution on the interval $[-3, 3]$. The factor $X_t = (x_{1t}, x_{2t}, x_{3t})^\tau$ was

$$\begin{cases} x_{1t} - \frac{3t}{n} = 0.8(x_{1t-1} - \frac{3t}{n}) + \varepsilon_{1t}, \ \ \varepsilon_{1t} = \sigma_{1t} e_{1t}, \ \sigma_{1t}^2 = 1.0 + 0.3\varepsilon_{1t-1}^2, \\ x_{2t} - \frac{t}{n} = -0.5(x_{2t-1} - \frac{t}{n}) + \varepsilon_{2t}, \ \varepsilon_{2t} = \sigma_{2t} e_{2t}, \ \sigma_{2t}^2 = 0.9 + 0.15\varepsilon_{2t-1}^2 + 0.7\sigma_{2t-1}^2, \\ x_{3t} + \frac{2t}{n} = 0.3(x_{3t-1} + \frac{2t}{n}) + \varepsilon_{3t}, \ \ \varepsilon_{3t} = \sigma_{3t} e_{3t}, \ \sigma_{3t}^2 = 1.1 + 0.2\varepsilon_{3t-1}^2 + 0.6\sigma_{3t-1}^2, \end{cases} \tag{3.2}$$

where e_{1t}, e_{2t}, and e_{3t} are independent $N(0, 1)$, and $\{x_{it}, i = 1, 2, 3\}$ are AR(1) processes with nonconstant mean.

We used $n = 50, 100, 200, 500,$ 1,000, and $d = 0.2n, 0.6n, 1.2n, 2n, 4n$, respectively. In each setting of the simulation, 200 replications were conducted.

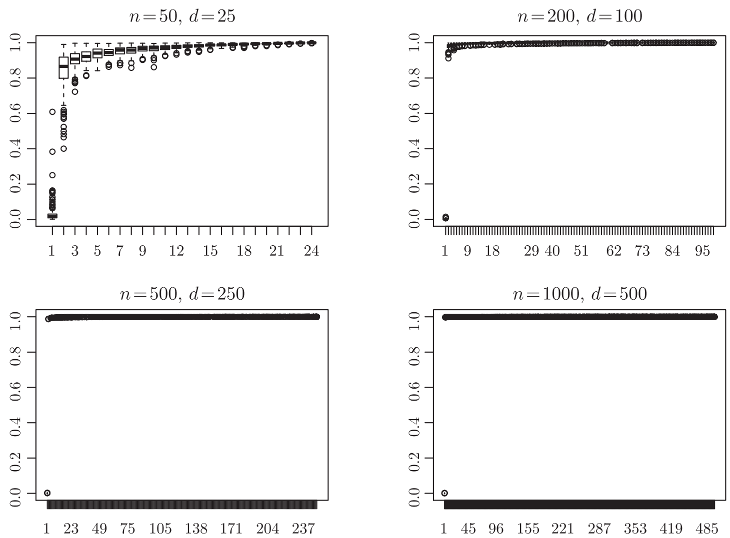

Figure 4. Boxplots for the ridge-type ratios $(\hat{\lambda}_{i+1} + \log(n)/10n)/(\hat{\lambda}_i + \log(n)/10n)$, $i = 1, \ldots, d$, for model (3.1).

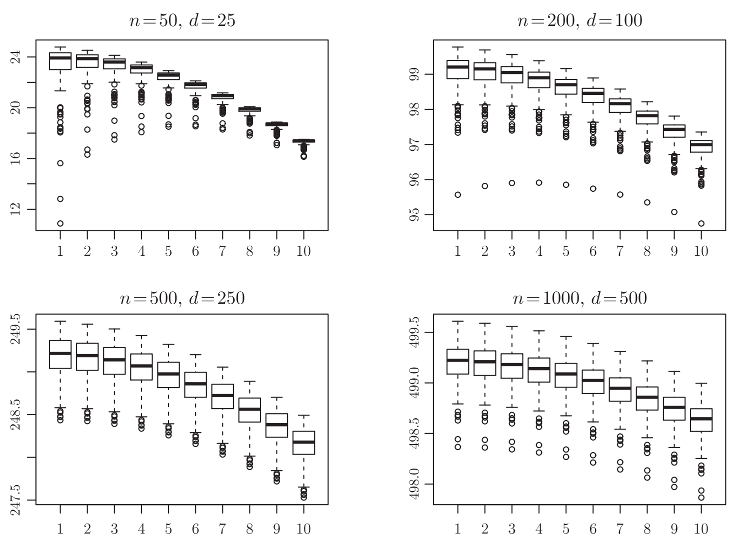

Figure 5. Boxplots for the BIC-type $G(i), i = 1, \ldots, 10$ for model (3.1).

The relative frequency estimates of the probability $P(\hat{r} = 3)$ are reported in Table 2, and the boxplots of the ratios $\hat{\lambda}_{i+1}/\hat{\lambda}_i$, the ridge-type ratios $(\hat{\lambda}_{i+1} + \log(n)/10n)/(\hat{\lambda}_i + \log(n)/10n)$, and the BIC-type estimates with $C_n = log(n)$ are, respectively, depicted in Figures 6, 7, and 8. Here, we chose $R = \min\{d/2, n/2\}$ for the ratio-based method to avoid using the last bad estimates of $\hat{\lambda}_{i+1}/\hat{\lambda}_i$. From Table 2, we have the following findings. First, when the sample size is

Table 2. Example 2: Relative frequency estimates for $P(\hat{r}=3)$ with 200 replicated samples in the simulation.

n	d	RE	RRE	BIC
50	$0.2n$	0.81	0.79	0.33
	$0.6n$	0.80	0.785	0.61
	$1.2n$	0.735	0.73	0.7
	$2n$	0.81	0.79	0.83
	$4n$	0.81	0.81	0.87
100	$0.2n$	0.95	0.93	0.75
	$0.6n$	0.97	0.965	0.935
	$1.2n$	0.94	0.945	0.97
	$2n$	0.95	0.94	0.985
	$4n$	0.93	0.93	0.99
200	$0.2n$	0.995	0.995	0.98
	$0.6n$	0.99	0.985	0.985
	$1.2n$	1	1	1
	$2n$	1	1	1
	$4n$	1	1	1
500	$0.2n$	1	1	1
	$0.6n$	1	1	1
	$1.2n$	1	1	1
	$2n$	1	1	1
	$4n$	1	1	1
1000	$0.2n$	1	1	1
	$0.6n$	1	1	1
	$1.2n$	1	1	1
	$2n$	1	1	1
	$4n$	1	1	1

small and the dimension is also small, the ratio estimate works best and both the ratio-based estimates much outperform the BIC-type estimate. See the results with $n=50,100$, and $d \leq 1.2n$. The ratio-based estimates are suitably used in case of small size of sample and dimension, whereas the BIC-type estimate can be applied otherwise. Second, the blessing of dimensionality is obvious for the BIC-type estimate for r, but it seems not clear for the ratio estimates. Figure 6 shows that the ratio-based methods seem not useful when we do not restrict R, especially with $d=2n$ and $4n$. This is due to the fact that the ratio around at $n-1$ is much smaller than the ratio at 3. Figures 7 and 8 indicate that the number of factors are estimated correctly by the ridge-type ratio and BIC-type estimates. Through this example, it seems that the phenomenon of “the blessing of dimensionality” (Lam and Yao (2012)) may not be true for the ratio and ridge-type ratio estimates, while may be true for the BIC-type estimate in our nonstationary setting of time series.

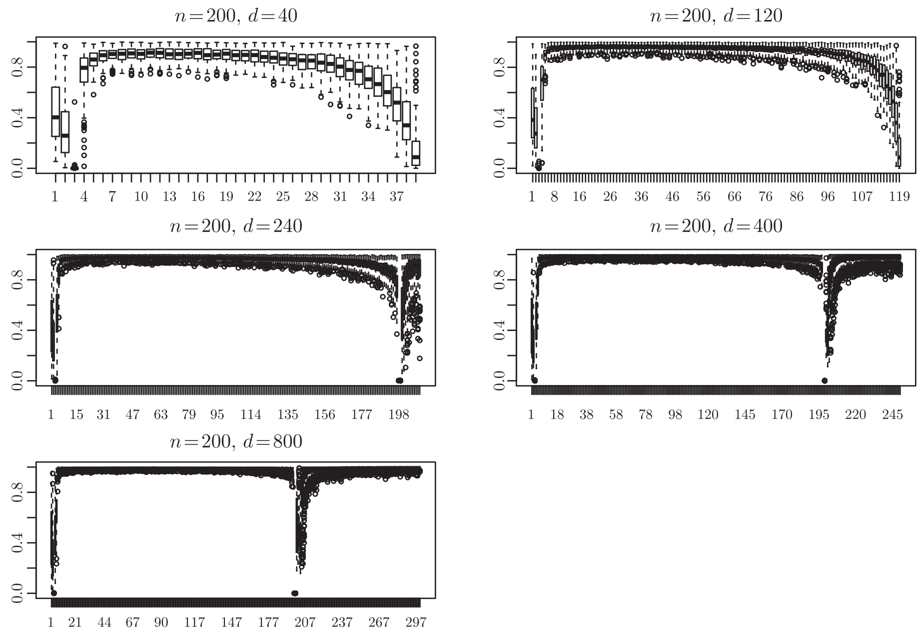

Figure 6. Boxplots for the ratios $\hat{\lambda}_{i+1}/\hat{\lambda}_i$ for model (3.2).

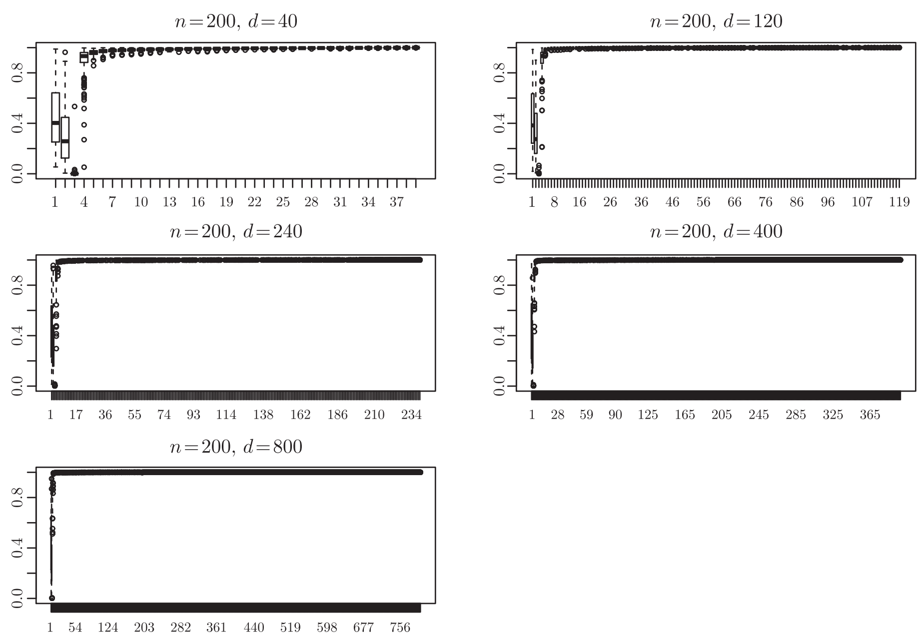

Figure 7. Boxplots for the the ridge-type ratios $(\hat{\lambda}_{i+1} + \log(n)/10n)/(\hat{\lambda}_i + \log(n)/10n)$ for model (3.2).

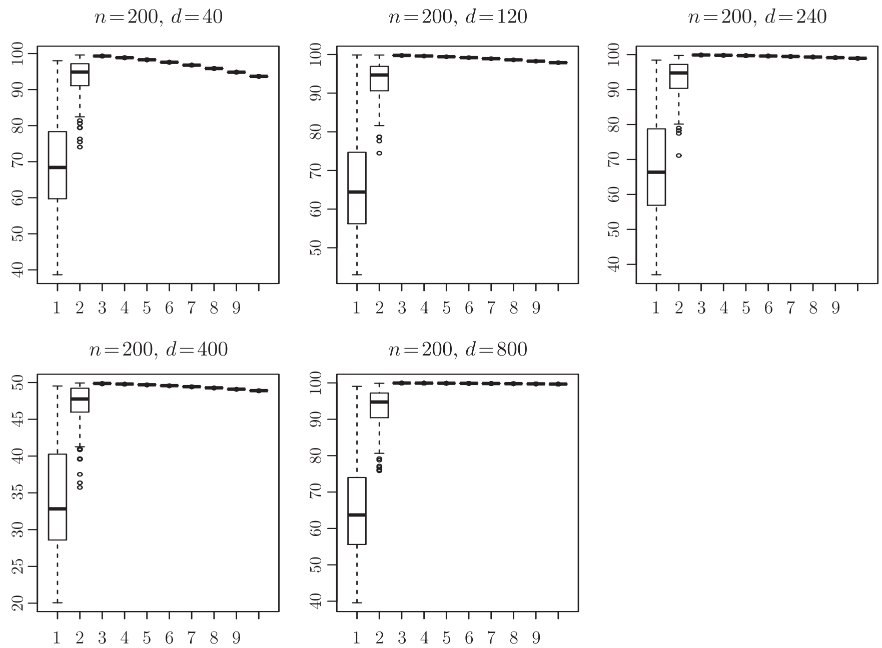

Figure 8. Boxplots for the BIC-type $G(i), i = 1, \ldots, 10$ for model (3.2).

To further investigate "the dimensionality blessing", another simulation was conducted by altering the above model. The second and third factors here differed from (3.2) through having a random walk and a deterministic trend. This example was based on Example 2 of Pan and Yao (2008), adding the volatility equations.

Example 3. For the model

$$\begin{cases} x_{1t} - \frac{3t}{n} = 0.8(x_{1t-1} - \frac{3t}{n}) + \varepsilon_{1t}, \quad \varepsilon_{1t} = \sigma_{1t}e_{1t}, \quad \sigma_{1t}^2 = 1.0 + 0.3\varepsilon_{1t-1}^2; \\ x_{2t} = x_{2t-1} + \sqrt{\frac{10}{n}}\varepsilon_{2t}, \quad \varepsilon_{2t} = \sigma_{2t}e_{2t}, \quad \sigma_{2t}^2 = 0.9 + 0.15\varepsilon_{2t-1}^2 + 0.7\sigma_{2t-1}^2; \\ x_{3t} = -\frac{2t}{n} + \varepsilon_{3t}, \quad \varepsilon_{3t} = \sigma_{3t}e_{3t}, \quad \sigma_{3t}^2 = 1.1 + 0.2\varepsilon_{3t-1}^2 + 0.6\sigma_{3t-1}^2, \end{cases} \tag{3.3}$$

with the simulation settings the same as in Example 2, we report the values of relative frequency estimates for the probability $P(\hat{r} = 3)$ in Table 3, and depict the boxplots of the ratio, the ridge-type ratio, and the BIC-type estimates in Figures 9, 10, and 11, respectively. In Table 3, the ratio estimate shows better than the ridge-type ratio estimate and again, when n and d are small, both outperform the BIC-type estimate. The dimensionality may not play a positive role for the ratio and ridge-type ratio estimates, whereas it still does for the BIC-type estimate. The boxplots of the ratio and ridge-type ratio estimates indicate

Table 3. Example 4: Relative frequency estimates for $P(\hat{r} = 3)$ with 200 replicated samples in the simulation.

n	d	RE	RRE	BIC
	$0.2n$	0.42	0.39	0.035
	$0.6n$	0.35	0.35	0.2
50	$1.2n$	0.375	0.36	0.24
	$2n$	0.375	0.355	0.38
	$4n$	0.365	0.35	0.59
	$0.2n$	0.39	0.35	0.12
	$0.6n$	0.39	0.38	0.37
100	$1.2n$	0.44	0.42	0.435
	$2n$	0.4	0.38	0.57
	$4n$	0.365	0.355	0.71
	$0.2n$	0.395	0.375	0.23
	$0.6n$	0.395	0.35	0.395
200	$1.2n$	0.34	0.355	0.53
	$2n$	0.36	0.34	0.645
	$4n$	0.4	0.375	0.8
	$0.2n$	0.34	0.28	0.36
	$0.6n$	0.355	0.3	0.54
500	$1.2n$	0.355	0.33	0.725
	$2n$	0.38	0.335	0.8
	$4n$	0.35	0.32	0.815
	$0.2n$	0.40	0.36	0.565
	$0.6n$	0.38	0.3	0.75
1000	$1.2n$	0.38	0.36	0.78
	$2n$	0.32	0.3	0.81
	$4n$	0.345	0.33	0.83

that the factor number r cannot be well estimated, because the boxplot of $\hat{\lambda}_3/\hat{\lambda}_2$ looks better than $\hat{\lambda}_4/\hat{\lambda}_3$. This may have resulted from the behavior of factor, since $\Sigma_X(1)$ is almost a singular matrix.

The ratio estimate works much as does the ridge-type ratio estimate, although the latter is consistent and, in Example 1, is better. The BIC-type estimate performs not as well as the ratio-based methods when the sample size n and the dimension d are small, while it works better for large n and d. Unlike in stationary time series cases, the ratio estimate does not have the dimensionality blessing property, whereas, the BIC-type estimate enjoys this property.

3.2. Data examples

Data set 1. The data are daily close prices: the daily log-returns of 46 stocks (times 100), which are component stocks of HSI in the period from 30 Nov.

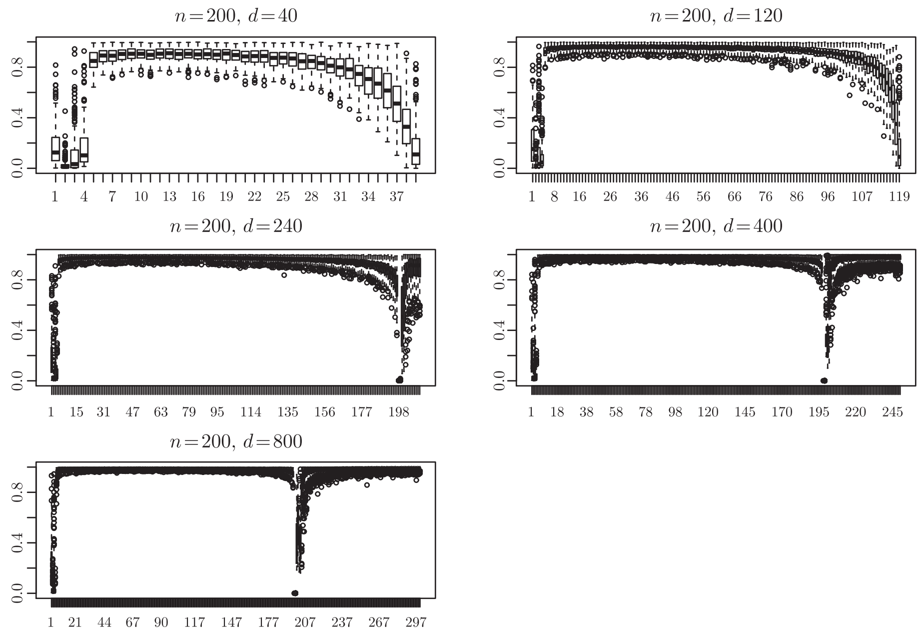

Figure 9. Boxplots for the ratios $\hat{\lambda}_{i+1}/\hat{\lambda}_i$ for model (3.3).

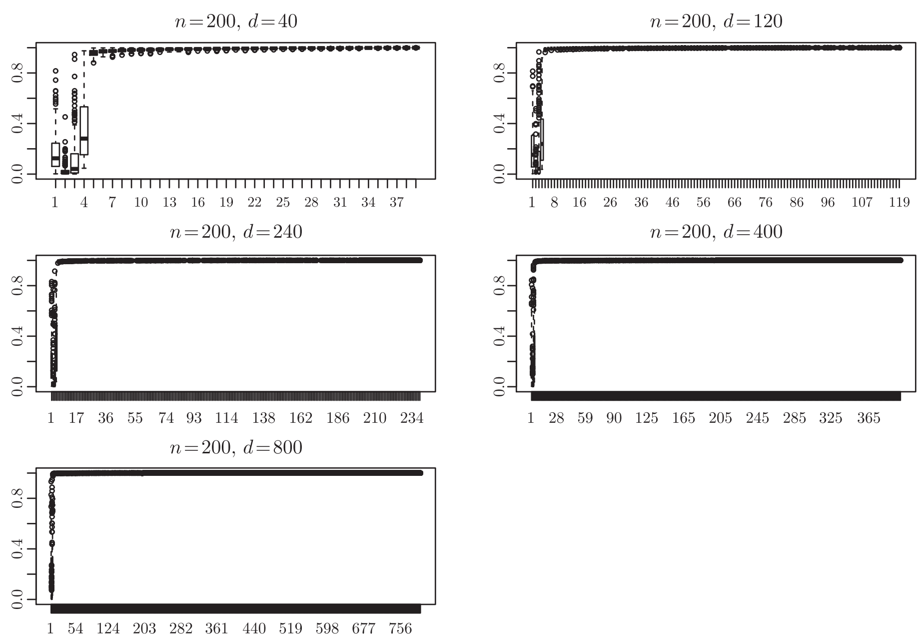

Figure 10. Boxplots for the the ridge-type ratios $(\hat{\lambda}_{i+1} + \log(n)/10n)/(\hat{\lambda}_i + \log(n)/10n)$ for model (3.3).

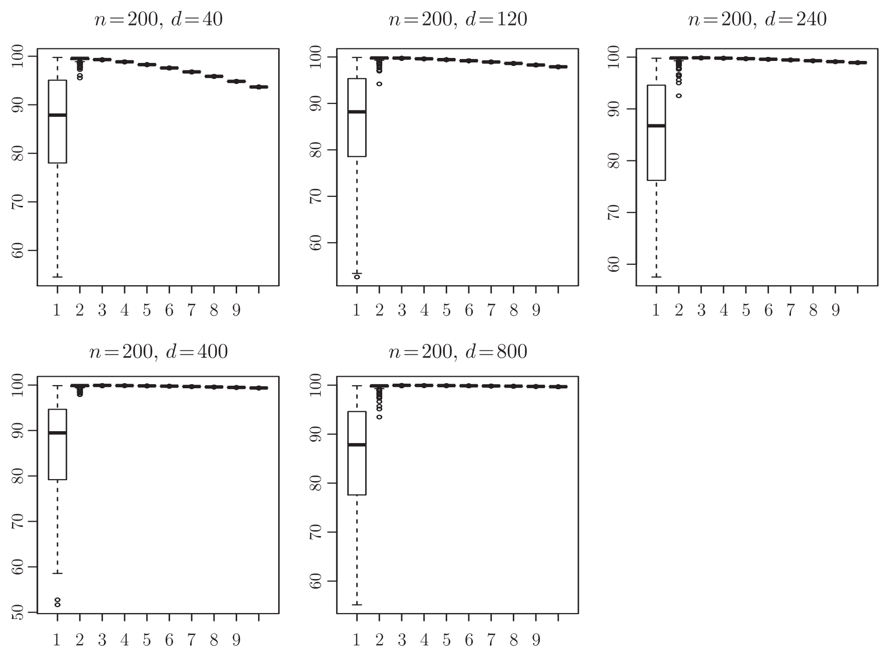

Figure 11. Boxplots for the BIC-type $G(i), i = 1, \ldots, 10$ for model (3.3).

2009 to 23 Nov. 2012. The data are obtained from *Yahoo finance*. In this data set, $n = 735$ and $d = 46$. As $d < n$, we used the ratio-based estimates to determine the number of factors. First, the eigenanalysis of the matrix $\hat{\Omega}$ as (2.4) was implemented. The eigenvalues and their ratios are plotted in Figure 12. From this figure, the $\hat{\lambda}_i$ are close to 0 for all $i \geq 3$, the ridge-type ratios $(\hat{\lambda}_{i+1} + \log(n)/10n)/(\hat{\lambda}_i + \log(n)/10n)$ are close to 1 for all $i \geq 4$, whereas the values of the ratios $\hat{\lambda}_{i+1}/\hat{\lambda}_i$ do not have a pattern so obvious. Both the ratio and the ridge-type ratio estimate $\hat{r}$ are 1. Meanwhile, the estimated orthogonal factor matrix $\hat{A}_{\hat{r}}$ and its orthogonal complement $\hat{B}_{\hat{r}}$ are easily obtained; their dimensions are, respectively, $\hat{r}$ and $d - \hat{r}$. For the estimated factor, Figure 13 suggests that its (ts)-plot is similar to the (ts)-plot of log-returns of HSI times 100 in the same time period. This also indicates that one factor model is sufficient for this data set.

In this data set, the heteroscedasticity seems to exist for each element of Y_t. Figure 14 presents the squared and the absolute correlograms about the first component of Y_t. For the estimated factor $Z_t = \hat{X}_t = \hat{A}^{\tau} Y_t$, we also depict its squared and absolute correlograms in Figure 14. It indicates the existence of heteroscedasticity in Z_t. When Z_t was fitted by a GARCH(1,1) model, the results were obtained

$$\hat{\Sigma}_{Z_t} = 0.9131 + 0.0750 Z_{t-1}^2 + 0.8984 \hat{\Sigma}_{Z_{t-1}} \tag{3.4}$$

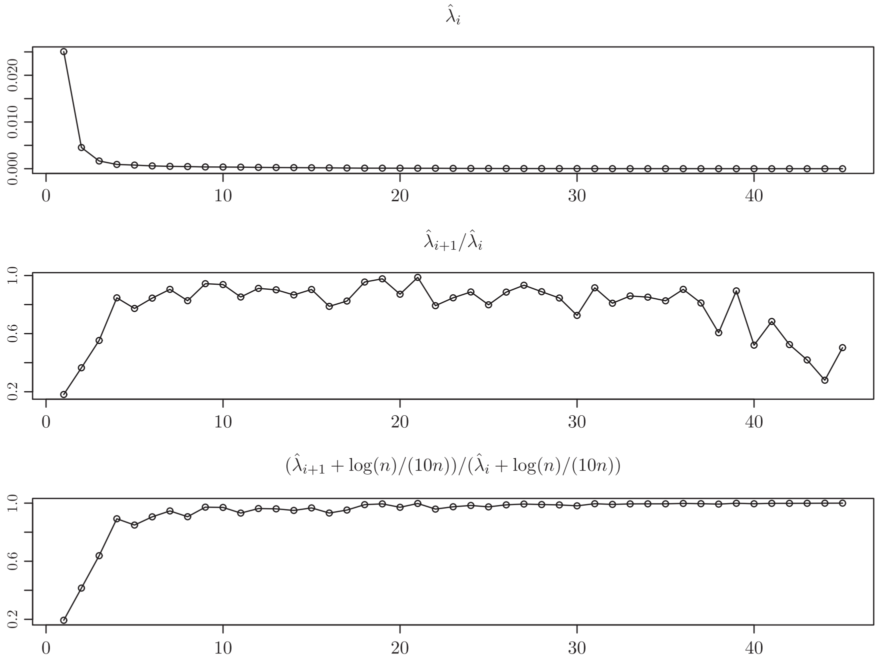

Figure 12. Plots of the estimated eigenvalues, the ratios and the ridge-type ratios.

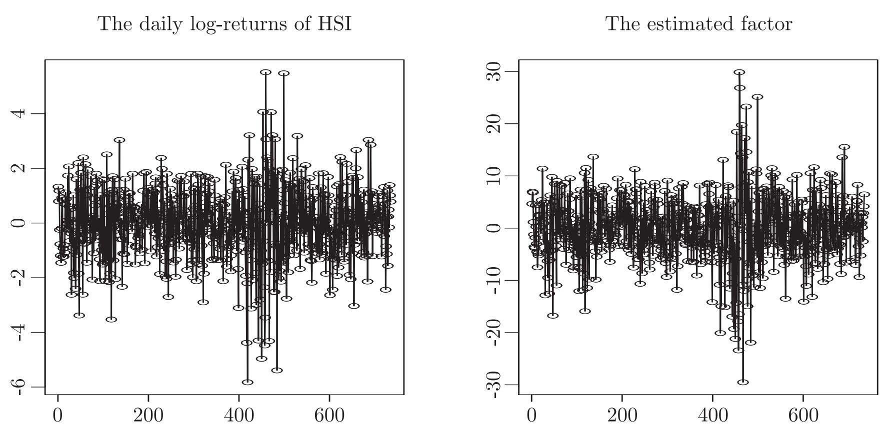

Figure 13. The time series plots of the estimated factors and the return series of the HSI index.

and an estimate $\hat{\Sigma}_t = \hat{A}_{\hat{r}}\hat{\Sigma}_{Z_t}\hat{A}_{\hat{r}}^{\tau} + \hat{A}_{\hat{r}}\hat{A}_{\hat{r}}^{\tau}\hat{\Sigma}_y\hat{B}_{\hat{r}}\hat{B}_{\hat{r}}^{\tau} + \hat{B}_{\hat{r}}\hat{B}_{\hat{r}}^{\tau}\hat{\Sigma}_y$ can be given.

Here, (2.1) implies that $\hat{X}_t = \hat{A}_{\hat{r}}^{\tau}Y_t$ and $\hat{\xi}_t = Y_t - \hat{A}_{\hat{r}}\hat{X}_t = Y_t - \hat{A}_{\hat{r}}\hat{A}_{\hat{r}}^{\tau}Y_t$. The residual correlograms of this model shows that the autocorrelation and partial autocorrelation functions of the residuals are almost all within their two standard error limits, suggesting that there is no significant residual serial correlation.

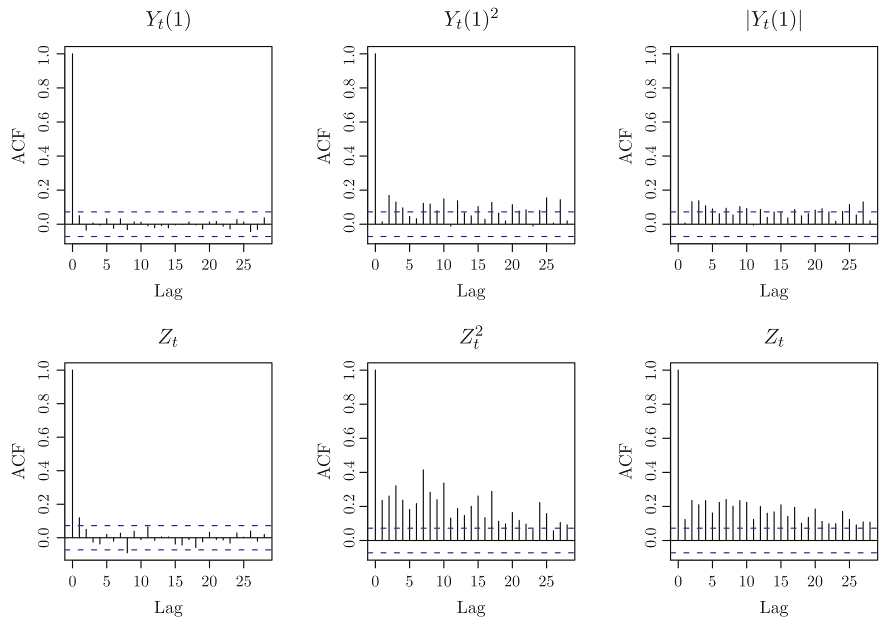

Figure 14. The correlograms about the behaviors of heteroscedasticity for Y_t and the estimated factor Z_t.

For this data set, modelling the volatility of multivariate time series is displyed through a lower dimensional factor process, that suggest that both the ratio and the ridge-type ratio are useful.

Data set 2. The macroeconomic data analysed in Eichler, Motta, and Von Sachs (2011) was constructed by a balanced panel of 22 monthly US macroeconomic time series from January 1960 to December 2003, which consists of 13 industrial production (IP) variables and 9 interest Rate (IR) variables. After the original data were transformed, Eichler, Motta, and Von Sachs (2011) found that the data were non-stationary with a deterministic time variation. For this data, n=528 and d=22. We first computed the eigenvalues of the matrix $\hat{\Omega}$ at (2.4), then plotted them, the ratios, the ridge-type ratios, and BIC in Figure 15.

Here we find that the eigenvalues are all very quite small, so the ridge-type ratios may be influenced as the ridge value would dominate the values in the ratios. Both the ratios and the ridge-type ratios suggest a one factor model. It appears that the number of factors that is determined by either the ratios or the ridge-type ratios tends to be small. However, Eichler, Motta, and Von Sachs (2011) made a careful analysis and found a one factor model wanting. In contrast, the BIC method suggests a three factor modelling, which coincides with

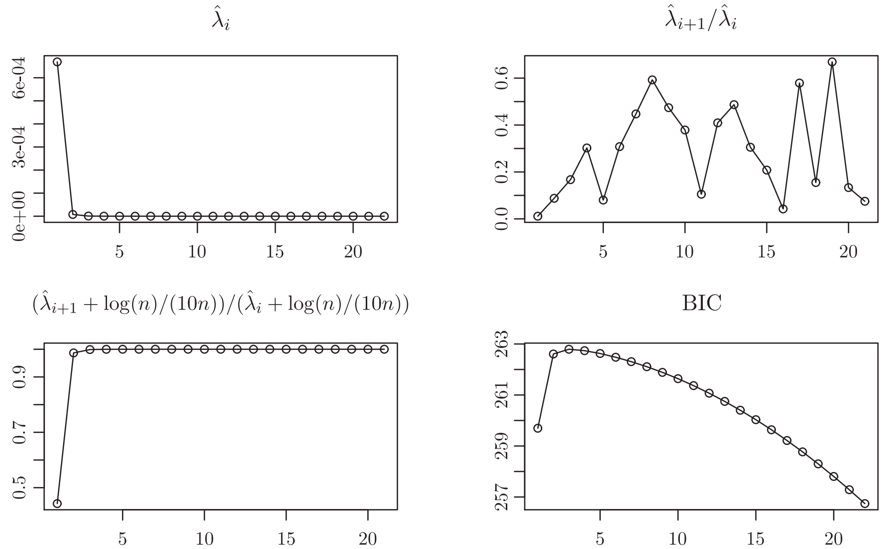

Figure 15. Plots of the estimated eigenvalues, the ratios, the ridge-type ratios and BIC.

the selection of Eichler, Motta, and Von Sachs (2011).

4. Conclusions

The aim of our work is to handle dimensionality determination in nonstationary and conditional heteroscedasticity settings. A ridge-type ratio and a BIC-type estimate are proposed and proved to be consistent. A comparison with the ratio estimate is conducted. The numerical studies suggest that, unlike stationary and conditional homoscedasticity settings, the dimensionality blessing may not be for the ratio-based methods to enjoy, while may still be for the BIC-type method. On the other hand, in practice, the ratio estimate, though not consistent, works well, particularly when the sample size and the dimension are small. From the limited simulations we conducted, a suggestion could be that, when $d \leq 1.2n$ and $n \leq 200$, use the ratio-based estimate; otherwise, use the BIC-type estimate. Further, to have an estimate sharing the advantages of both the ratio-based estimate and the BIC-type estimate, we might develop a hybrid of them by a data-driven manner. A study of this in ongoing.

Acknowledgement

The authors thank the editor, associate editor, and a referee for their constructive suggestions and comments that led to a significant improvement of an early version of the manuscript. Qiang Xia gratefully acknowledges the support of K.C.Wong Education Foundation, Hong Kong. He worked at College of Science, South China Agricultural University, he was also supported by the National Social Science Foundation of China (No:12CTJ019) and a Project of Humanities and Social Sciences (Project No.11YJCZH195) funded by Ministry of Education of China. Wangli Xu's research was supported by the National Natural Science Foundation of China (No. 11071253) and Beijing Nova Programme (2010B066). Lixing Zhu's research was supported by a grant from the University Grants Council of Hong Kong, China.

References

Anderson, T. W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika* **28**, 1-25.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71**, 135-171.

Cook, R. D. (1998). *Regression Graphics.* Wiley.

Eichler, M., Motta, G. and Von Sachs, R. (2011). Fitting dynamic factor models to non-stationary time series. *J. Econometrics* **163**, 51-70.

Engle, R. F. (2002). Dynamic conditional correlation a simple class of multivariate GARCH models. *J. Business and Economic Statistics* **20**, 339-350.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalised ARCH. *Econom. Theory* **11**, 122-150.

Fan, J., Wang, M. and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *J. Roy. Statist. Soc. Ser. B* **70**, 679-702.

Feng, Z. H., Wen, X. R., Yu, Z. and Zhu, L. X. (2013). On partial sufficient dimension reduction with applications to partially linear multi-index models. *J. Amer. Statist. Assoc.* **108**, 237-246.

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.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004). The generalized dynamic-factor model: consistency and rates. *Journal of Econometrics* **119**, 231-255.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic-factor model: one-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100**, 830-840.

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

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.* **40**, 694-726.

Pan, J., Peña D., Polonik, W. and Yao, Q. (2011). Modelling multivariate volatilities via common factors. Available at `http://stats.lse.ac.uk/q.yao/qyao.links/paper/pppy.pdf`.

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

Peña, D. and Box, E. P. (1987). Identifying a simplifying structure in time series. *J. Amer. Statist. Assoc.* **82**, 836-843.

Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *J. Statist. Plann. Inference* **136**, 1237-1257.

Priestley, M. B., Rao, T. S. and Tong, H. (1974). Applications of principal component analysis and factor analysis in the identification of multivariable systems. *IEEE Trans. Automat. Control* **19**, 703-704.

Wang, H. (2012). Factor profiling for ultra high dimensional variable selection. *Biometrika* **99**, 15-28.

Zhu, L. P. and Zhu, L. X. (2009). On distribution weighted partial least squares with diverging number of highly correlated predictors. *J. Roy. Statist. Soc. Ser. B* **71**, 525-548.

Zhu, L. X., Miao, B. Q. and Peng, H. (2006). On sliced inverse regression with high dimensional covariates. *J. Amer. Statist. Assoc.* **101**, 630-643.

School of Statistics, Renmin University of China, Beijing, 100872, China.

E-mail: xiaqiang@scau.edu.cn

School of Statistics, Renmin University of China, Beijing, 100872, China.

E-mail: wxu.stat@gmail.com

The Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China.

E-mail: lzhu@hkbu.edu.hk

(Received August 2013; accepted May 2014)