Large Volatility Matrix Inference via Combining Low-Frequency and High-Frequency Approaches

Minjing TAO, Yazhen WANG, Qiwei YAO, and Jian ZOU

It is increasingly important in financial economics to estimate volatilities of asset returns. However, most of the available methods are not directly applicable when the number of assets involved is large, due to the lack of accuracy in estimating high-dimensional matrices. Therefore it is pertinent to reduce the effective size of volatility matrices in order to produce adequate estimates and forecasts. Furthermore, since high-frequency financial data for different assets are typically not recorded at the same time points, conventional dimension-reduction techniques are not directly applicable. To overcome those difficulties we explore a novel approach that combines high-frequency volatility matrix estimation together with low-frequency dynamic models. The proposed methodology consists of three steps: (i) estimate daily realized covolatility matrices directly based on high-frequency data, (ii) fit a matrix factor model to the estimated daily covolatility matrices, and (iii) fit a vector autoregressive model to the estimated volatility factors. We establish the asymptotic theory for the proposed methodology in the framework that allows sample size, number of assets, and number of days go to infinity together. Our theory shows that the relevant eigenvalues and eigenvectors can be consistently estimated. We illustrate the methodology with the high-frequency price data on several hundreds of stocks traded in Shenzhen and Shanghai Stock Exchanges over a period of 177 days in 2003. Our approach pools together the strengths of modeling and estimation at both intra-daily (high-frequency) and inter-daily (low-frequency) levels.

KEY WORDS: Dimension reduction; Eigenanalysis; Factor model; Matrix process; Realized volatilities; Vector autoregressive model.

1. INTRODUCTION

Modeling and forecasting the volatilities of financial returns are vibrant research areas in econometrics and statistics. For financial data at daily or longer time horizons, which are often referred to as low-frequency data, there exists extensive literature on direct volatility modeling using GARCH, discrete stochastic volatility, and diffusive stochastic volatility models as well as indirect modeling using implied volatility obtained from option pricing models. See Wang (2002).

With the availability of intraday financial data, which are called high-frequency data, there is an surging interest on estimating volatilities using high-frequency returns directly. The field of high-frequency finance has experienced a rapid evolvement in past several years. One of the focus points at present is to estimate integrated volatility over a period of time, say, a day. Estimation methods for univariate volatilities include realized volatility (RV), bi-power realized variation (BPRV), two-time scale realized volatility (TSRV), wavelet realized volatility (WRV), realized kernel volatility (KRV), preaveraging realized volatility, and Fourier realized volatility (FRV). For the cases with multiple assets, a so-called nonsynchronized problem arises, which refers to the fact that transactions for different assets often occur at distinct times, and the high-frequency prices of different assets are recorded at mismatched time points. Hayashi and Yoshida (2005) and Zhang (2011) proposed to estimate integrated covolatility of the two assets based on overlap intervals and previous ticks, respectively. Barndorff-Nielsen et al. (2010) employed a refresh time scheme to synchronize the data and then applied a realized kernel to the synchronized data for estimating integrated covolatility. Christensen, Kinnebrock, and Podolskij (2010) studied integrated covolatility estimation by the preaveraging approach. Nevertheless most existing works on volatility estimation using high-frequency data are for a single asset or a small number of assets, and therefore are only directly applicable when the integrated volatility concerned is either a scalar or a small matrix.

In reality we often face with scenarios involving a large number of assets. The integrated volatility concerned then is a matrix of a large size. In principle, a large volatility matrix may be estimated as follows: estimating each diagonal element; representing an integrated volatility of a single asset; by univariate methods such as RV and BPRV; and estimating each off-diagonal element, representing an integrated covolatility of two assets, by the method of Hayashi and Yoshida (2005) or Zhang (2011). However, due to the large number of elements in the volatility matrix, such a naive estimator often behaves poorly. It is widely known that as dimension (or matrix size) goes to infinity, the estimators such as sample covariance matrix and usual realized covolatility estimators are inconsistent in the sense that the eigenvalues and eigenvectors of the matrix estimators are far from the true targets (Johnstone 2001; Johnstone and Lu 2009; and Wang and Zou 2010). Banding and tresholding are proposed by Bickel and Levina (2008a, 2008b) to yield consistent estimators of large covariance matrices, and a factor model approach is used in Fan, Fan, and Lv (2008) to estimate large covariance matrices. To illustrate this point, we conduct a simulation as follows: consider p assets over unit time interval with all log prices following independent standard Brownian motions. Observations were taken without noise at the same time grids $t_i = i/n$ for $i = 0, 1, \ldots, n$. Then the true integrated volatility

Minjing Tao is Ph.D. Student, and Yazhen Wang is Professor (E-mail: yzwang@stat.wisc.edu), Department of Statistics, University of Wisconsin–Madison, 1300 University Avenue, Madison, WI 53706. Qiwei Yao is Professor, Department of Statistics, London School of Economics, Houghton Street, London, WC2A 2AE, U.K. and Guanghua School of Management, Peking University, Beijing. Jian Zou is Assistant Professor, Department of Mathematical Sciences, Indiana University–Purdue University, 402 North Blackford Street, Indianapolis, IN 46202. Yazhen Wang's research was partially supported by the NSF grant DMS-105635. Qiwei Yao was partially supported by the EPSRC grants EP/C549058/1 and EP/H010408/1. The authors thank the editor, associate editor, and two anonymous referees for stimulating comments and suggestions, which led to significant improvements in both substance and the presentation of the article.

Journal of the American Statistical Association
September 2011, Vol. 106, No. 495, Theory and Methods
DOI: 10.1198/jasa.2011.tm10276

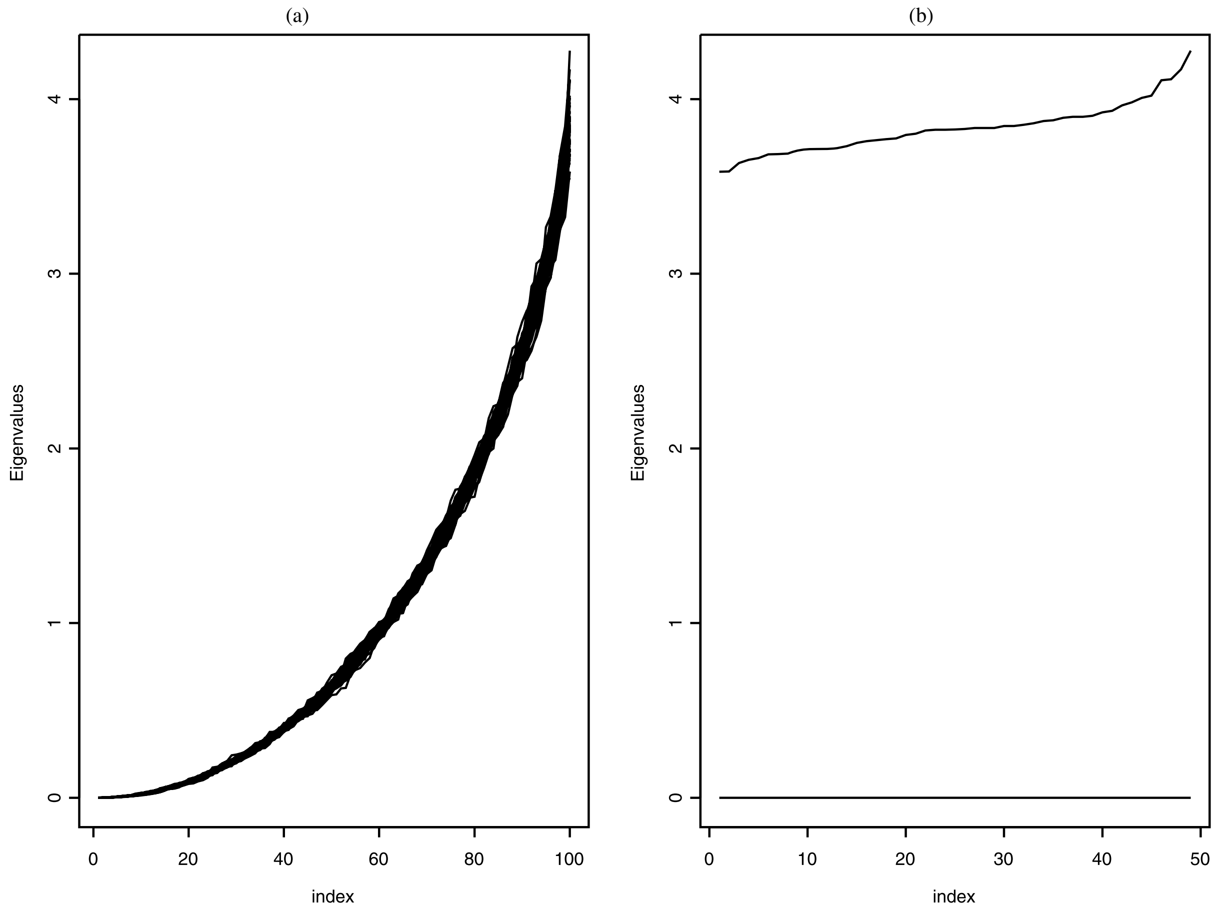

Figure 1. Plots of eigenvalues $\widehat{\mathbf{V}}$ from a simulation with 50 repetitions. (a) Each of the 50 curves represents the ordered 100 eigenvalues of each sampled $\widehat{\mathbf{V}}$. (b) The minimum and maximum eigenvalues of $\widehat{\mathbf{V}}$ across 50 repetitions.

matrix $\mathbf{V}$ is the identity matrix $\mathbf{I}_p$. The estimator for $\mathbf{V}$ based on the RV and the co-RV methods is

$$\widehat{\mathbf{V}} = (\widehat{V}_{jk}), \quad \text{with } \widehat{V}_{jk} = \frac{1}{n}\sum_{i=1}^{n} Z_{ij} Z_{ik} \text{ for } 1 \le j, k \le p,$$

where Z_{ij}, $i = 1, \ldots, n, j = 1, \ldots, p$, are effectively independent N(0, 1) random variables. Setting $p = 100$, we drew 50 samples of size $n = 100$. For each of 50 samples, we computed the 100 eigenvalues of $\widehat{\mathbf{V}}$ and evaluated their maximum and minimum eigenvalues. Of the 50 sets of 100 eigenvalues, we found that all sets range approximately from zero to four with an average minimum eigenvalue 0.0001 and an average maximum eigenvalue 3.9. This clearly indicates the serious lack of accuracy in estimating $\mathbf{V}$ since all its eigenvalues are equal to 1. The inaccuracy of the estimator $\widehat{\mathbf{V}}$ is further manifested by the wide range of its eigenvalues displayed in Figure 1. This numerical experiment indicates that it is essential to reduce the number of estimated parameters in such a high-dimensional problem.

This article considers high-frequency prices observed on a large number of assets over many days. We propose a matrix factor model for daily integrated volatility matrix processes. The matrix factor model facilitates combining high-frequency volatility estimation with low-frequency dynamic models as well as reducing an effective dimension in large volatility matrices. It is important to note that the proposed matrix factor model is directly for integrated volatility matrices. Since prices for different assets are typically observed at different times, it is often impossible to apply an ordinary factor model to the original price data directly. Nevertheless the available abundance of the information in high-frequency data should make modeling daily volatilities easier. Indeed the inference for our matrix factor model is more direct than that for the ordinary factor volatility models for price data.

Our estimation procedure consists of three steps. First we estimate integrated volatility matrix for each day by thresholding average realized volatility matrix (TARVM) estimators. We then perform an eigenanalysis to fit a matrix factor model for the estimated daily integrated volatility matrices and obtain estimated daily factor matrices. Finally we fit a vector autoregressive (VAR) model for the estimated daily volatility factor matrices. The proposed methodology pools together strengths in modeling and estimation at both low-frequency and high-frequency levels. In the univariate case where dimension reduction is not an issue, Andersen, Bollerslev, and Diebold (2003) and Corsi (2009) demonstrated that the forecasting for volatilities may be improved from fitting a heterogeneous AR model to RV and BPRV based estimators of integrated volatilities. The approach is termed as the HAR–RV model. Our proposal may be viewed as a high-dimensional version of the HAR–RV approach based on new idea on matrix factor modeling.

We have established novel asymptotic theory for the proposed methodology in the framework that allows p (number of assets), n (average sample size), and L (number of days) all go to infinity. The established convergence rates for TARVM estimators and the matrix factor model under matrix norm provide a theoretical justification for the proposed methodology. These results indicate that the relevant eigenvalues and eigenvectors in the proposed factor modeling can be consistently estimated for large p. We also show that fitting the VAR model with the estimated daily volatility factor matrices from high-frequency data is asymptotically as efficient as that with true daily volatility factor matrices.

The rest of the article is organized as follows. The proposed methodology is presented in Section 2. Its asymptotic theory is established in Section 3. Numerical illustration is reported in Section 4. Section 5 features conclusions. All proofs are collected in the Appendix.

2. METHODOLOGY

2.1 Price Model and Observed Data

Suppose that there are p assets and their log price process $\mathbf{X}(t) = \{X_1(t), \ldots, X_p(t)\}^T$ obeys an Itô process governed by

$$d\mathbf{X}(t) = \boldsymbol{\mu}_t\, dt + \boldsymbol{\sigma}_t\, d\mathbf{W}_t, \qquad t \in [0, L], \tag{1}$$

where L is an integer, $\mathbf{W}_t$ is a p-dimensional standard Brownian motion, $\boldsymbol{\mu}_t$ is a drift taking values in $\mathbb{R}^p$, and $\boldsymbol{\sigma}_t$ is a $p \times p$ matrix. Both $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are assumed to be continuous in t. Let a day be a unit time. The integrated volatility matrix for the ℓth day is defined as

$$\boldsymbol{\Sigma}_x(\ell) = \int_{\ell-1}^{\ell} \boldsymbol{\sigma}_s \boldsymbol{\sigma}_s^T\, ds, \qquad \ell = 1, \ldots, L.$$

Suppose that high-frequency prices for the ith asset on the ℓth day are observed at times $t_{ij} \in (\ell - 1, \ell]$, $\ell = 1, \ldots, L$. We denote by $Y_i(t_{ij})$ the observed log price of the ith asset at time t_{ij}. Due to the so-called nonsynchronized problem, typically $t_{i_1 j} \neq t_{i_2 j}$ for any $i_1 \neq i_2$. Furthermore the high-frequency prices are typically masked by some microstructure noise in the sense that the observed log price $Y_i(t_{ij})$ is a noisy version of the corresponding true log price $X_i(t_{ij})$. A common practice is to assume

$$Y_i(t_{ij}) = X_i(t_{ij}) + \varepsilon_i(t_{ij}), \tag{2}$$

where $\varepsilon_i(t_{ij})$ are iid noise with mean zero and variance η_i, and $\varepsilon_i(\cdot)$ and $X_i(\cdot)$ are independent with each other.

Let $n_i(\ell)$ be the sample size for asset i on the ℓth day, that is, $n_i(\ell)$ = the number of $t_{ij} \in (\ell - 1, \ell]$, $n(\ell) = \sum_{i=1}^{p} n_i(\ell)/p$, the average sample size of the p assets on the ℓth day, and $n = \sum_{\ell=1}^{L} n(\ell)/L$, the average sample size across the p assets and over all L days.

2.2 Realized Volatility Matrix Estimator

To highlight the basic idea in realized volatility matrix estimation, we first consider estimating $\boldsymbol{\Sigma}_x(1)$, the integrated volatility matrix on day one, by averaging realized volatility matrix (ARVM) estimator proposed in Wang and Zou (2010). Suppose that $\boldsymbol{\tau} = \{\tau_r, r = 1, \ldots, m\}$ is a predetermined sampling frequency. For asset i, define previous-tick times

$$\tau_{i,r} = \max\{t_{ij} \le \tau_r, j = 1, \ldots, n_i(1)\}, \qquad r = 1, \ldots, m.$$

Based on $\boldsymbol{\tau}$ we define realized covolatility between assets i_1 and i_2 by

$$\tilde{\Sigma}_y(1, \boldsymbol{\tau})[i_1, i_2] = \sum_{r=1}^{m} \left[Y_{i_1}\left(\tau_{i_1,r}\right) - Y_{i_1}\left(\tau_{i_1,r-1}\right)\right] \times \left[Y_{i_2}\left(\tau_{i_2,r}\right) - Y_{i_2}\left(\tau_{i_2,r-1}\right)\right], \tag{3}$$

and realized volatility matrix by

$$\tilde{\boldsymbol{\Sigma}}_y(1, \boldsymbol{\tau}) = (\tilde{\Sigma}_y(1, \boldsymbol{\tau})[i_1, i_2])_{1 \le i_1, i_2 \le p}. \tag{4}$$

We take the predetermined sampling frequency $\boldsymbol{\tau}$ as the following regular grids. Given a fixed m, there are $K = [n(1)/m]$ classes of nonoverlap regular grids given by

$$\begin{aligned}\boldsymbol{\tau}^k &= \{(r-1)/m, r = 1, \ldots, m\} + (k-1)/n(1) \\ &= \{(r-1)/m + (k-1)/n(1), r = 1, \ldots, m\},\end{aligned} \tag{5}$$

where $k = 1, \ldots, K$, and $n(1)$ is the average sample size on day one. For each $\boldsymbol{\tau}^k$, using (3) and (4) we define realized covolatility $\tilde{\Sigma}_y(1, \boldsymbol{\tau}^k)[i_1, i_2]$ between assets i_1 and i_2 and realized volatility matrix $\tilde{\boldsymbol{\Sigma}}_y(1, \boldsymbol{\tau}^k)$. The ARVM estimator is given by

$$\tilde{\Sigma}_y(1)[i_1, i_2] = \frac{1}{K} \sum_{k=1}^{K} \tilde{\Sigma}_y(1, \boldsymbol{\tau}^k)[i_1, i_2] - 2m\widehat{\eta}_{i_1} 1(i_1 = i_2), \tag{6}$$

$$\tilde{\boldsymbol{\Sigma}}_y(1) = (\tilde{\Sigma}_y(1)[i_1, i_2]) = \frac{1}{K} \sum_{k=1}^{K} \tilde{\boldsymbol{\Sigma}}_y(1, \boldsymbol{\tau}^k) - 2m\widehat{\boldsymbol{\eta}}, \tag{7}$$

where

$$\widehat{\eta}_i = \frac{1}{2n_i(1)} \sum_{j=1}^{n_i(1)} [Y_i(t_{i,j}) - Y_i(t_{i,j-1})]^2, \tag{8}$$

are estimators of noise variances η_i, and $\widehat{\boldsymbol{\eta}} = \mathrm{diag}(\widehat{\eta}_1, \ldots, \widehat{\eta}_p)$ is the estimator of $\boldsymbol{\eta} = \mathrm{diag}(\eta_1, \ldots, \eta_p)$. The averaging in (6) and (7) is to reduce the impact of microstructure noise on realized volatility matrices $\tilde{\boldsymbol{\Sigma}}_y(1, \boldsymbol{\tau}^k)$ and yield a better ARVM estimator.

When p is small, $\tilde{\boldsymbol{\Sigma}}_y(1)$ provides a good estimator for $\boldsymbol{\Sigma}_x(1)$. But for large p, it is well known that $\tilde{\boldsymbol{\Sigma}}_y(1)$ is inconsistent. In fact, statistics theory for small n and large p or large n but much larger p problems shows that the eigenvalues and the eigenvectors of, for example, a sample covariance matrix or a realized volatility matrix are inconsistent estimators for the corresponding true eigenvalues and eigenvectors. The proposed methodology in this article relies on consistent estimation of eigenvalues and eigenvectors of large volatility matrices. To estimate $\boldsymbol{\Sigma}_x(1)$ consistently for large p, we need impose some sparsity structure on $\boldsymbol{\Sigma}_x(1)$ [see (18) in Section 3] and threshold $\tilde{\boldsymbol{\Sigma}}_y(1)$ by retaining its elements whose absolute values exceed a given value and replacing others by zero. See Bickel and Levina (2008a, 2008b), Johnstone and Lu (2009), Wang and Zou (2010). We threshold $\tilde{\boldsymbol{\Sigma}}_y(1)$ and obtain an estimator

$$\widehat{\boldsymbol{\Sigma}}_y(1) = \mathcal{T}_{\varpi}[\tilde{\boldsymbol{\Sigma}}_y(1)] = \left(\tilde{\Sigma}_y(1)[i_1, i_2] 1_{(|\tilde{\Sigma}_y[i_1, i_2]| \ge \varpi)}\right), \tag{9}$$

where ϖ is a threshold. The (i_1, i_2)th element of $\widehat{\boldsymbol{\Sigma}}_y(1)$ is equal to $\tilde{\Sigma}_y(1)[i_1, i_2]$ if its absolute value is greater than or equal to ϖ and zero otherwise. The threshold ARVM estimator $\widehat{\boldsymbol{\Sigma}}_y(1)$ is called TARVM estimator.

Similarly, based on high-frequency data on the ℓth day we construct ARVM estimator $\tilde{\boldsymbol{\Sigma}}_y(\ell)$ and define TARVM estimator $\widehat{\boldsymbol{\Sigma}}_y(\ell)$ to provide an estimator for the integrated volatility matrix $\boldsymbol{\Sigma}_x(\ell)$, $\ell = 2, \ldots, L$.

2.3 A Matrix Factor Model

To reduce the effective number of entries in $\boldsymbol{\Sigma}_x(\ell)$ and connect high-frequency volatility matrix estimation with low-frequency volatility dynamic models, we propose a factor model as follows:

$$\boldsymbol{\Sigma}_x(\ell) = \mathbf{A}\boldsymbol{\Sigma}_f(\ell)\mathbf{A}^T + \boldsymbol{\Sigma}_0, \qquad \ell = 1, \ldots, L, \tag{10}$$

where r is a fixed small integer (much smaller than p), $\boldsymbol{\Sigma}_0$ is a $p \times p$ positive definite constant matrix, $\boldsymbol{\Sigma}_f(\ell)$ are $r \times r$ positive definite matrices and treated as factor volatility process, and $\mathbf{A}$ is a $p \times r$ factor loading matrix. This effectively assumes that the daily dynamical structure of the matrix process $\boldsymbol{\Sigma}_x(\ell)$ is driven by that of a lower-dimensional latent process $\boldsymbol{\Sigma}_f(\ell)$, while $\boldsymbol{\Sigma}_0$ represents the static part of $\boldsymbol{\Sigma}_x(\ell)$. Although the form of the above model is similar to the factor volatility models proposed by, for example, Engle, Ng, and Rothschild (1990), the key difference here is that we have the "observations" $\widehat{\boldsymbol{\Sigma}}_y(\cdot)$ directly on the volatility process $\boldsymbol{\Sigma}_x(\cdot)$. Since the high-frequency prices are measured at the different times for different assets, we cannot apply a factor model directly to the observed high-frequency data $Y_i(t_{ij})$.

The availability of the estimators for $\boldsymbol{\Sigma}_x(\cdot)$ from high-frequency data makes it easier to estimate both the factor loading matrix $\mathbf{A}$ and the factor volatility $\boldsymbol{\Sigma}_f(\cdot)$. In fact the estimation problem now reduces to a standard eigenanalysis and can be easily performed for p as large as a few thousands. This is in marked contrast to the more standard circumstances when only the observations on $\mathbf{X}_t$ are available; see, for example, Pan and Yao (2008). To fix the idea, let us temporarily assume that we observe $\boldsymbol{\Sigma}_x(\ell)$. Note that there is no loss of generality in assuming $\mathbf{A}$ in (10) satisfying the condition $\mathbf{A}^T\mathbf{A} = \mathbf{I}_r$. In fact, $\mathbf{A}$ is still not completely identifiable even under this constraint, however the linear space spanned by the columns of $\mathbf{A}$ is. Note that there exists a $p \times (p - r)$ matrix $\mathbf{B}$ for which $\mathbf{B}^T\mathbf{A} = \mathbf{0}$ and $\mathbf{B}^T\mathbf{B} = \mathbf{I}_{p-r}$, that is, $(\mathbf{A}, \mathbf{B})$ is a $p \times p$ orthogonal matrix. Now multiplying $\mathbf{B}^T$ on both sides of (10), we obtain that

$$\mathbf{B}^T\boldsymbol{\Sigma}_x(\ell) = \mathbf{B}^T\boldsymbol{\Sigma}_0. \tag{11}$$

Put

$$\bar{\boldsymbol{\Sigma}}_x = \frac{1}{L}\sum_{\ell=1}^{L}\boldsymbol{\Sigma}_x(\ell), \qquad \bar{\mathcal{S}}_x = \frac{1}{L}\sum_{\ell=1}^{L}\{\boldsymbol{\Sigma}_x(\ell) - \bar{\boldsymbol{\Sigma}}_x\}^2. \tag{12}$$

Equation (11) implies that for all $\ell = 1, \ldots, L$, $\mathbf{B}^T\boldsymbol{\Sigma}_x(\ell) = \mathbf{B}^T\bar{\boldsymbol{\Sigma}}_x$, and

$$\begin{aligned}\mathbf{B}^T\bar{\mathcal{S}}_x\mathbf{B} &= \frac{1}{L}\sum_{\ell=1}^{L}\{\mathbf{B}^T\boldsymbol{\Sigma}_x(\ell) - \mathbf{B}^T\bar{\boldsymbol{\Sigma}}_x\}\{\boldsymbol{\Sigma}_x(\ell)\mathbf{B} - \bar{\boldsymbol{\Sigma}}_x\mathbf{B}\}\\ &= \mathbf{0}.\end{aligned} \tag{13}$$

This suggests that the columns of $\mathbf{B}$ are the $p - r$ orthonormal eigenvectors of $\bar{\mathcal{S}}_x$, corresponding to the $(p - r)$-fold eigenvalue 0. The other r orthonormal eigenvectors of $\bar{\mathcal{S}}_x$, corresponding to the r nonzero eigenvalues, may be taken as the columns of the factor loading matrix $\mathbf{A}$.

Of course $\boldsymbol{\Sigma}_x(\ell)$ is unknown in practice. We use $\widehat{\boldsymbol{\Sigma}}_y(\ell)$ as a proxy. Let

$$\bar{\boldsymbol{\Sigma}}_y = \frac{1}{L}\sum_{\ell=1}^{L}\widehat{\boldsymbol{\Sigma}}_y(\ell), \qquad \bar{\mathcal{S}}_y = \frac{1}{L}\sum_{\ell=1}^{L}\{\widehat{\boldsymbol{\Sigma}}_y(\ell) - \bar{\boldsymbol{\Sigma}}_y\}^2, \tag{14}$$

where $\widehat{\boldsymbol{\Sigma}}_y(\ell)$ are TARVM estimators computed from high-frequency data; see Section 2.2 above. Then the estimator $\widehat{\mathbf{A}}$ is obtained using the r orthonormal eigenvectors of $\bar{\mathcal{S}}_y$, corresponding to the r largest eigenvalues, as its columns. Consequently the estimated factor volatilities are

$$\widehat{\boldsymbol{\Sigma}}_f(\ell) = \widehat{\mathbf{A}}^T\widehat{\boldsymbol{\Sigma}}_y(\ell)\widehat{\mathbf{A}}, \qquad \ell = 1, \ldots, L, \tag{15}$$

and the estimator for $\boldsymbol{\Sigma}_0$ in model (10) may be taken as

$$\widehat{\boldsymbol{\Sigma}}_0 = \bar{\boldsymbol{\Sigma}}_y - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\bar{\boldsymbol{\Sigma}}_y\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T. \tag{16}$$

2.4 VAR Modeling for Factor Volatilities

With estimated factor volatility matrices in (15), we build up the dynamical structure of $\boldsymbol{\Sigma}_x(\ell)$ by fitting a VAR model to $\widehat{\boldsymbol{\Sigma}}_f(\ell)$. One alternative is to adopt more sophisticated multivariate volatility models to fit $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ or $\widehat{\boldsymbol{\Sigma}}_f^{1/2}(\ell)$ (see Wang and Yao 2005 and Remark A.1 after Lemma A.6 in the Appendix). We opt to a simple VAR model in the spirit of the HAR–RV approach advocated by Andersen, Bollerslev, and Diebold (2003) and Corsi (2009). They demonstrate that fitting an AR model to realized (one-dimensional) volatilities may lead to significant improvement in volatility forecasting.

For a $r \times r$ matrix $\boldsymbol{\Sigma}$, let $\mathrm{vech}(\boldsymbol{\Sigma})$ be the $r(r+1)/2 \times 1$ vector obtained by stacking together the truncated column vectors of $\boldsymbol{\Sigma}$, where the truncating means to remove all the elements above the main diagonal. Then the VAR model for $\boldsymbol{\Sigma}_f(\ell)$ is of the form

$$\mathrm{vech}\{\boldsymbol{\Sigma}_f(\ell)\} = \boldsymbol{\alpha}_0 + \sum_{j=1}^{q}\boldsymbol{\alpha}_j\,\mathrm{vech}\{\boldsymbol{\Sigma}_f(\ell - j)\} + \mathbf{e}_\ell, \tag{17}$$

where $q \geq 1$ is an integer, $\boldsymbol{\alpha}_0$ is a vector, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q$ are square matrices, and $\mathbf{e}_\ell$ is a vector white noise process with zero mean and finite fourth moments. Since $\boldsymbol{\Sigma}_f(\ell)$ are estimated by $\widehat{\boldsymbol{\Sigma}}_f(\ell)$, with a fixed q, we adopt the least squares estimators $\widehat{\boldsymbol{\alpha}}_j$ for the coefficients $\boldsymbol{\alpha}_j$, which are the minimizer of

$$\sum_{\ell=q+1}^{L}\left\|\mathrm{vech}\{\widehat{\boldsymbol{\Sigma}}_f(\ell)\} - \boldsymbol{\alpha}_0 - \sum_{j=1}^{q}\boldsymbol{\alpha}_i\,\mathrm{vech}\{\widehat{\boldsymbol{\Sigma}}_f(\ell - j)\}\right\|^2,$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector. The order q may be determined by, for example, the standard criteria such as AIC or BIC.

3. ASYMPTOTIC THEORY

First we introduce some notations. Given a p-dimensional vector $\mathbf{x} = (x_1, \ldots, x_p)^T$ and a p by p matrix $\mathbf{U} = (U_{ij})$, define matrix norm as follows,

$$\|\mathbf{U}\|_2 = \sup\{\|\mathbf{U}\mathbf{x}\|_2, \|\mathbf{x}\|_2 = 1\}, \qquad \|\mathbf{x}\|_2 = \left(\sum_{i=1}^{p}|x_i|^2\right)^{1/2}.$$

Then $\|\mathbf{U}\|_2$ is equal to the square root of the largest eigenvalue of $\mathbf{U}^T\mathbf{U}$, where $\mathbf{U}^T$ is the transpose of $\mathbf{U}$, and for symmetric $\mathbf{U}$, $\|\mathbf{U}\|_2$ is equal to its largest absolute eigenvalue.

Second we state the following assumptions for the asymptotic analysis.

(A1) We assume all row vectors of $\mathbf{A}^T$ and $\mathbf{\Sigma}_0$ in factor model (10) obey the sparsity condition (18) below. For a p-dimensional vector $\mathbf{x} = (x_1, \ldots, x_p)^T$, we say it is sparse if it satisfies

$$\sum_{i=1}^{p} |x_i|^\delta \leq C\pi(p), \tag{18}$$

where $\delta \in [0, 1)$, C is a positive constant, and $\pi(p)$ is a deterministic function of p that grows slowly in p with typical examples $\pi(p) = 1$ or $\log p$.

(A2) Assume factor model (10) has fixed r factors, with $\mathbf{A}^T\mathbf{A} = I_r$, and matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_f$ in (10) satisfy

$$\|\mathbf{\Sigma}_0\|_2 < \infty, \qquad \max_{1\leq \ell \leq L} |\Sigma_f(\ell)[j,j]| = O_P(\log L), \qquad j = 1, \ldots, r.$$

(A3) We impose the following moment conditions on diffusion drift $\boldsymbol{\mu}_t = (\mu_1(t), \ldots, \mu_p(t))^T$ and diffusion variance $\boldsymbol{\sigma}_t = (\sigma_{ij}(t))_{1\leq i,j\leq p}$ in price model (1) and microstructure noise $\varepsilon_i(t_{ij})$ in data model (2): for some $\beta \geq 4$,

$$\max_{1\leq i\leq p} \max_{0\leq t\leq L} E\big[|\sigma_{ii}(t)|^\beta\big] < \infty,$$

$$\max_{1\leq i\leq p} \max_{0\leq t\leq L} E\big[|\mu_i(t)|^{2\beta}\big] < \infty,$$

$$\max_{1\leq i\leq p} \max_{0\leq t_{ij}\leq L} E\big[|\varepsilon_i(t_{ij})|^{2\beta}\big] < \infty.$$

(A4) Each of p assets has at least one observation between τ_r^k and τ_{r+1}^k. That is, in the construction of ARVM estimator we assume $m = o(n)$, and

$$C_1 \leq \min_{1\leq i\leq p} \min_{1\leq \ell\leq L} \frac{n_i(\ell)}{n} \leq \max_{1\leq i\leq p} \max_{1\leq \ell\leq L} \frac{n_i(\ell)}{n} \leq C_2,$$

$$\max_{1\leq i\leq p} \max_{1\leq \ell\leq L} \max_{1\leq j\leq n_i(\ell)} |t_{ij} - t_{i,j-1}| = O(n^{-1}).$$

(A5) The characteristic polynomial of VAR model (17) has no roots in the unit circle so that it is a casual VAR model.

Remark 1. Condition (A1) together with factor model (10) imply that $\mathbf{\Sigma}_x(\ell)$ are sparse, which is required to consistently estimate $\mathbf{\Sigma}_x(\ell)$ for large p and will be shown by Lemma A.2 in the Appendix. When $\delta = 0$ in (18), sparsity refers to that there are at most $C\pi(p)$ number of nonzero coordinates in $\mathbf{x} = (x_1, \ldots, x_p)^T$, and matrix sparsity means that each row has at most $C\pi(p)$ number of nonzero elements. Sparsity is often a reasonable assumption for large volatility matrices. We may further improve sparsity for the volatility matrices by transformations such as removing the overall market effect and the sector effect. Condition (A2) imposes realistic bounded eigenvalues on $\mathbf{\Sigma}_0$ and a logarithm temporal growth on $\mathbf{\Sigma}_f(\ell)$ over $[0, L]$. As $\mathbf{\Sigma}_0$ is a constant matrix and $\mathbf{\Sigma}_f(\ell)$ are small matrices of fixed size r, Condition (A2) together with factor model (10) guarantee that the maximum eigenvalue of $\mathbf{\Sigma}_x(\ell)$ is free of p and has only order $\log L$, which will be proved in Lemma A.1 in the Appendix. The logarithm rate in (A2) is rather weak and reasonable, as the maxima of sequences of independent and typically dependent random variables are of a logarithm order. The assumption is to relieve from specifying temporal and cross-section dependence structures on the volatilities over time and across assets. Condition (A3) is the minimal moment requirements for the price process and microstructure noise. Condition (A4) is a technical condition that ensures adequate number of observations between grids and establishes the asymptotic theory for the proposed methodology. Condition (A5) is a standard condition for stationary AR time series.

We establish the asymptotic theory for the proposed models and the associated estimation methods. Since p, n, and L stand for dimension (number of assets), average daily observations, and the number of days, we let p, n, and L all go to infinity in the asymptotics. The two theorems below give the eigenvalue and eigenvector convergence for the difference between $\bar{\mathcal{S}}_x$ and $\bar{\mathcal{S}}_y$ defined in (12) and (14), respectively.

Theorem 1. Suppose models (1), (2) and (10) satisfy Conditions (A1)–(A4). As n, p, L all go to infinity, we have

$$\|\bar{\mathcal{S}}_y - \bar{\mathcal{S}}_x\|_2 = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big),$$

where $e_n \sim n^{-1/6}$ for the noise case and $e_n \sim n^{-1/3}$ for the no noise case [i.e., $\varepsilon_i(t_{ij}) = 0$ in (2)], and threshold ϖ used in (9) is of order $e_n(p^2L)^{1/\beta}\log L$.

Theorem 2. Suppose models (1), (2), and (10) satisfy Conditions (A1)–(A4). Denote the ordered eigenvalues of $\bar{\mathcal{S}}_x$ by $\lambda_1 \geq \cdots \geq \lambda_p$. Assume that there is a positive constant c such that $\lambda_j - \lambda_{j+1} \geq c$ for $j = 1, \ldots, r$. Let $\mathbf{a}_1, \ldots, \mathbf{a}_r$ be the eigenvectors of $\bar{\mathcal{S}}_x$ corresponding to the r largest eigenvalues $\lambda_1, \ldots, \lambda_r$. Also set $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_r$ be the r largest eigenvalues of $\bar{\mathcal{S}}_y$ and $\widehat{\mathbf{a}}_1, \ldots, \widehat{\mathbf{a}}_r$ the corresponding eigenvectors. Let $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_r)$ and $\widehat{\mathbf{A}} = (\widehat{\mathbf{a}}_1, \ldots, \widehat{\mathbf{a}}_r)$. Then as n, p, L go to infinity, we have

$$\mathbf{A}^T\widehat{\mathbf{A}} - \mathbf{I}_r = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big),$$

$$\widehat{\mathbf{\Sigma}}_f(\ell) - \mathbf{\Sigma}_f - \mathbf{A}^T\mathbf{\Sigma}_0\mathbf{A} = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big),$$

where e_n and ϖ are the same as in Theorem 1, and since the matrices are of fixed size r, the convergence holds under any matrix norms.

Remark 2. Since $e_n(p^2L)^{1/\beta}$ is powers of n, p, L while $\pi(p)\log^2 L$ depends on p and L through logarithm and thus is negligible in comparison with $[e_n(p^2L)^{1/\beta}]^{1-\delta}$. So the convergence rate is nearly equal to $[e_n(p^2L)^{1/\beta}]^{1-\delta}$. To consistently estimate the r largest eigenvalues and their corresponding eigenvectors of $\bar{\mathcal{S}}_x$ we need to make $e_n(p^2L)^{1/\beta}$ go to zero. As $e_n \sim n^{-1/3}$ for the noiseless case and $n \sim n^{-1/6}$ for the noise case, $e_n(p^2L)^{1/\beta}$ goes to zero if p^2L grows more slowly than $n^{\beta/3}$ for the noiseless case and $n^{\beta/6}$ for the noise case. For reasonably large β in moment Assumption (A3), the consistent requirement can accommodate the scenario when p is comparable to or larger than n. Thus, Theorems 1 and 2 establish the valid theoretical foundation for the proposed methodology in the sense that it yields consistent estimators of the r largest eigenvalues and their corresponding eigenvectors for the factor-based analysis under the large p scenario.

Next we establish asymptotic theory for parameter estimation in the VAR model (17) based on high-frequency data.

Theorem 3. Suppose that $\widehat{\boldsymbol{\alpha}}_i$ are least squares estimators of $\boldsymbol{\alpha}_i$ based on data $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ from the VAR model (17) and we denote by $\tilde{\boldsymbol{\alpha}}_i$ the least squares estimators of $\boldsymbol{\alpha}_i$ based on oracle data $\boldsymbol{\Sigma}_f(\ell)$ from the same VAR model (17). Then under Conditions (A1)–(A5) and the eigenvalue assumption of Theorem 2,

$$\widehat{\boldsymbol{\alpha}}_0 - \tilde{\boldsymbol{\alpha}}_0 - \mathrm{vech}\{\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}\} = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big),$$

$$\widehat{\boldsymbol{\alpha}}_i - \tilde{\boldsymbol{\alpha}}_i = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big), \qquad i = 1, \dots, q.$$

In particular, as $n, p, L \to \infty$, if $\pi(p)[e_n(p^2L)^{1/\beta}]^{1-\delta}L^{1/2} \times \log^2 L \to 0$, then

$$L^{1/2}\{\widehat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0 - \mathrm{vech}(\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}), \widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1, \dots, \widehat{\boldsymbol{\alpha}}_q - \boldsymbol{\alpha}_q\}$$

has the same limiting distribution as $L^{1/2}(\tilde{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0, \tilde{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1, \dots, \tilde{\boldsymbol{\alpha}}_q - \boldsymbol{\alpha}_q)$.

Remark 3. Theorem 3 shows that the proposed data-driven method of model fitting based on $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ estimated from high-frequency data can asymptotically achieve the same result as an oracle that uses true $\boldsymbol{\Sigma}_f(\ell)$ for model fitting. In other words, fitting the VAR model with the estimated daily volatility factor matrices from high-frequency data can be asymptotically as efficient as that with true daily volatility factor matrices.

Remark 4. We may replace the ARVM estimator used in the first stage by other volatility matrix estimators, for example, in Barndorff-Nielsen et al. (2008, 2010), Christensen, Kinnebrock, and Podolskij (2010), Griffin and Oomen (2011), Hautsch, Kyj, and Oomen (2009), and Zhang (2011). However, these estimators enjoy good properties only for the fixed matrix size p that is very small relative to sample size. When p is allowed to grow with sample size and its magnitude is comparable to sample size, all the estimators become inconsistent. Regularization adjustment such as thresholding is needed to make them consistent. For example, to improve the convergence rate of the ARVM estimator we may use the multiscale scheme in Fan and Wang (2007, section 4.3) and Zhang (2006) to construct the following multiscale realized volatility matrix (MRVM) estimator,

$$\tilde{\boldsymbol{\Sigma}}_y^*(1) = \sum_{m=1}^{\kappa} a_m\widehat{\boldsymbol{\Gamma}}^{K_m} + \zeta(\widehat{\boldsymbol{\Gamma}}^{K_1} - \widehat{\boldsymbol{\Gamma}}^{K_\kappa}),$$

where κ is the integer part of $\sqrt{n}$, $\widehat{\boldsymbol{\Gamma}}^{K_m}$ is defined via (3) and (4) as follows:

$$\widehat{\boldsymbol{\Gamma}}^{K_m} = \frac{1}{K_m}\sum_{k=1}^{K_m}\tilde{\boldsymbol{\Sigma}}_y(1, \boldsymbol{\tau}^k) = \left(\frac{1}{K_m}\sum_{k=1}^{K_m}\tilde{\Sigma}_y(1, \boldsymbol{\tau}^k)[i_1, i_2]\right)_{1\le i_1, i_2\le p},$$

$$K_m = m + \kappa, \qquad a_m = \frac{12(m+\kappa)(m-\kappa/2-1/2)}{\kappa(\kappa^2-1)},$$

$$\zeta = \frac{(2\kappa)(\kappa+1)}{(n+1)(\kappa-1)}.$$

For fixed p and noisy data, the ARVM estimator $\tilde{\boldsymbol{\Sigma}}_y(1)$ in (7) has convergence rate $n^{-1/6}$, while the MRVM estimator $\tilde{\boldsymbol{\Sigma}}_y^*(1)$ can achieve the optimal convergence rate $n^{-1/4}$ (Tao, Wang, and Chen 2011). However, as p goes to infinity and p and n are comparable, $\tilde{\boldsymbol{\Sigma}}_y^*(1)$ becomes inconsistent. Similar to (9) we need to threshold $\tilde{\boldsymbol{\Sigma}}_y^*(1)$ and obtain

$$\widehat{\boldsymbol{\Sigma}}_y^*(1) = \mathcal{T}_\varpi[\tilde{\boldsymbol{\Sigma}}_y^*(1)] = \big(\tilde{\Sigma}_y^*(1)[i_1, i_2]1_{(|\tilde{\Sigma}_y^*[i_1,i_2]|\ge\varpi)}\big),$$

where ϖ is a threshold. Similarly we can define $\widehat{\boldsymbol{\Sigma}}_y^*(\ell)$ for $\ell = 2, \dots, L$. If daily integrated volatility matrices $\boldsymbol{\Sigma}_x(\ell)$ are estimated by $\widehat{\boldsymbol{\Sigma}}_y^*(\ell)$ instead of $\widehat{\boldsymbol{\Sigma}}_y(\ell)$ for performing eigenanalysis and fitting the matrix factor and VAR models described in Sections 2.3 and 2.4, we expect to obtain the same conclusions as in Theorems 1–3 but with $e_n \sim n^{-1/4}$ for the noisy data case.

4. NUMERICAL EXAMPLES

We illustrate the proposed methodology with two sets of high-frequency data, the tick by tick prices of the 410 stocks traded in Shenzhen Stock Exchange and the 630 stocks traded in Shanghai Stock Exchange over a period of 177 days in 2003. The daily average intraday observations over the 177 days range from 194 to 1384 with overall average 578 for the stocks traded in the Shenzhen market and from 210 to 1620 with overall average 575 for the stocks traded in the Shanghai market.

4.1 Eigenanalysis Based on Estimated Daily Integrated Volatility Matrices

For each of the 177 days, we compute the estimated daily integrated volatility matrices using TARVM estimator in (9) with grids being selected in accord of 5 minute returns and thresholds being the top 5% of the largest absolute entries. This yields a sequence of 177 matrices of $\widehat{\boldsymbol{\Sigma}}_y(\ell)$, $\ell = 1, \dots, L = 177$, where the daily integrated volatility matrices for Shenzhen and Shanghai datasets are of sizes 410 by 410 and 630 by 630, respectively. The eigenvalues and eigenvectors of the sample variance matrix $\bar{\mathcal{S}}_y$ are then evaluated, and the 20 largest eigenvalues, multiplied by 1000, are plotted in Figures 2 and 3 for Shenzhen and Shanghai datasets, respectively. The plots show that the largest eigenvalue for the Shenzhen data and the two largest eigenvalues for the Shanghai data are much larger than the corresponding other eigenvalues, which are in a much smaller magnitude and decrease slowly.

4.2 A Simulation Study on Volatility Factor Selection

Theorems 1 and 2 imply that the eigenvalue difference between $\bar{\mathcal{S}}_y$ and $\bar{\mathcal{S}}_x$ converges in probability to zero, where $\bar{\mathcal{S}}_x$ has r positive eigenvalues and $p - r$ zero eigenvalues. Thus we may select r such that the smallest $p - r$ eigenvalues of $\bar{\mathcal{S}}_y$ are close to 0 while the r largest eigenvalues are significantly larger. Figures 2 and 3 suggest $r = 1$ and $r = 2$ for the datasets from the Shenzhen and Shanghai Exchanges, respectively. We conduct a simulation study below to provide some support for such empirical selection of r.

In the simulation study we consider two scenarios with $r = 1$ and $r = 2$, where $p = 410$ and $L = 177$. The simulation proceeds as follows. For the case of $r = 1$, we generate $\boldsymbol{\Sigma}_f(\ell)$ from an AR(1) model with mean, AR coefficient and noise variance being $(6, 0.65, 0.3)$ and then simulate $\boldsymbol{\Sigma}_x(\ell)$ from the matrix factor model (10) with loading matrix $\mathbf{A}$ formed by the

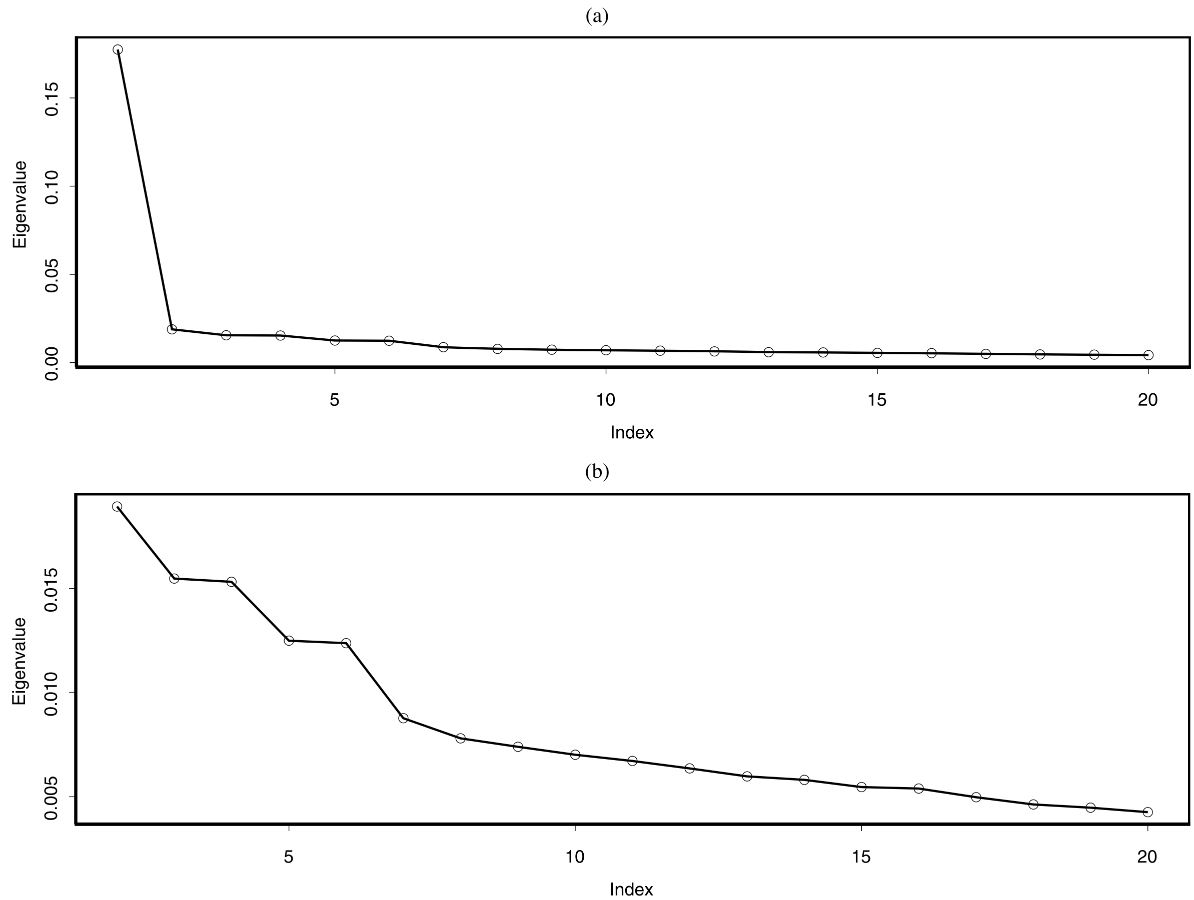

Figure 2. Plots of the 20 largest eigenvalues of $\bar{\mathcal{S}}_y$ for the dataset from Shenzhen Stock Exchange. (a) The plot of all 20 largest eigenvalues. (b) The plot of the second largest to 20th largest eigenvalues.

eigenvector corresponding to the largest eigenvalue of $\bar{\mathcal{S}}_y$ obtained from the Shenzhen data. For the case of $r = 2$, we take $\Sigma_f(\ell)[1, 2] = \Sigma_f(\ell)[2, 1] = 0$, and generate $\Sigma_f(\ell)[1, 1]$ and $\Sigma_f(\ell)[2, 2]$ from two AR(1) models with mean, AR coefficient and noise variance being $(6, 0.65, 0.3)$ and $(4, 0.5, 0.3)$, respectively, and we simulate $\mathbf{\Sigma}_x(\ell)$ from the matrix factor model (10) with loading matrix $\mathbf{A}$ formed by the two eigenvectors corresponding to the two largest eigenvalues of $\bar{\mathcal{S}}_y$ obtained from the Shenzhen data.

We simulate high-frequency price data from model (1) with zero drift by discretizing the diffusion equation

$$\mathbf{X}(t_k) = \mathbf{X}(t_{k-1}) + \boldsymbol{\sigma}_{t_{k-1}}\big[\mathbf{W}_{t_k} - \mathbf{W}_{t_{k-1}}\big],$$

where $t_k = \ell - 1 + k/n$, $k = 1, \ldots, n$, $n = 200$, $\ell = 1, \ldots, 177$, during the period of the ℓth day, we take σ_{t_k} to be $\mathbf{A}[\mathbf{\Sigma}_f(\ell) + 0.32\mathbf{Z}_k]^{1/2}\mathbf{A}^T$, $\mathbf{Z}_k = (Z_k[j_1, j_2])_{1\le j_1, j_2 \le r}$ are r by r matrices whose entries $Z_k[j_1, j_2]$ are standard normal random variables with temporal correlation $\mathrm{corr}(Z_k[j_1, j_2], Z_{k'}[j_1, j_2]) = \exp(-|k - k'|)$, and zero correlation for different entries, that is, $\mathrm{corr}(Z_k[j_1, j_2], Z_{k'}[j_1', j_2']) = 0$ for $(j_1, j_2) \neq (j_1', j_2')$. Finally, data $Y_i(t_k)$ are obtained from model (2) by adding to $\mathbf{X}(t_k)$ iid normal noise with mean zero and standard deviation 0.064. We calculate ARVM estimator $\tilde{\mathbf{\Sigma}}_y(\ell)$ based on the data in the ℓth day and the threshold estimator $\widehat{\mathbf{\Sigma}}_y(\ell)$ as described in Section 2.2. According to the description in Section 2.3 we compute $\bar{\mathcal{S}}_y$ from $\widehat{\mathbf{\Sigma}}_y(\ell)$ and then the eigenvalues and eigenvectors of $\bar{\mathcal{S}}_y$. We repeat the whole simulation procedure 100 times. As in Wang and Zou (2010), estimators $\widehat{\mathbf{\Sigma}}_y(\ell)$ are tuned to minimize its estimated mean squares error based on 100 repetitions. Figure 4 plots the 20 largest eigenvalues of $\bar{\mathcal{S}}_y$ over the 100 simulated samples for the cases of $r = 1$ and $r = 2$. The plots show that for the case of $r = 1$, the largest eigenvalues are clustered around 0.5, and for the case of $r = 2$, the two largest eigenvalues are fluctuated around 0.5 and 0.4, respectively, and these large eigenvalues are much larger than other eigenvalues in the corresponding cases, where these small eigenvalues are close to zero. Moreover, the clusters in Figure 4 for the 100 simulated samples are apparently quite tight and separate. The simulation results indicate that the largest eigenvalue and the two largest eigenvalues for the respective cases of $r = 1$ and $r = 2$ are significant and hence the selection of volatility factors based on large eigenvalues matches very well with the true values of r in the corresponding cases.

The daily average intraday observations over the 177 days for the stocks traded in the Shenzhen and Shanghai markets are from around 200 to over 1000. As the simulation results reported above are for the case with 200 intraday observations, we have tried to increase intraday observations from 200 to 600 and 1000 in the simulation study and found the similar cluster patterns for the eigenvalues. In fact, the eigenvalue clusters become tighter as the number of intraday observations increases.

The procedure in Hansen and Lunde (2006) is used to calculate the noise to signal ratios for the simulated and real

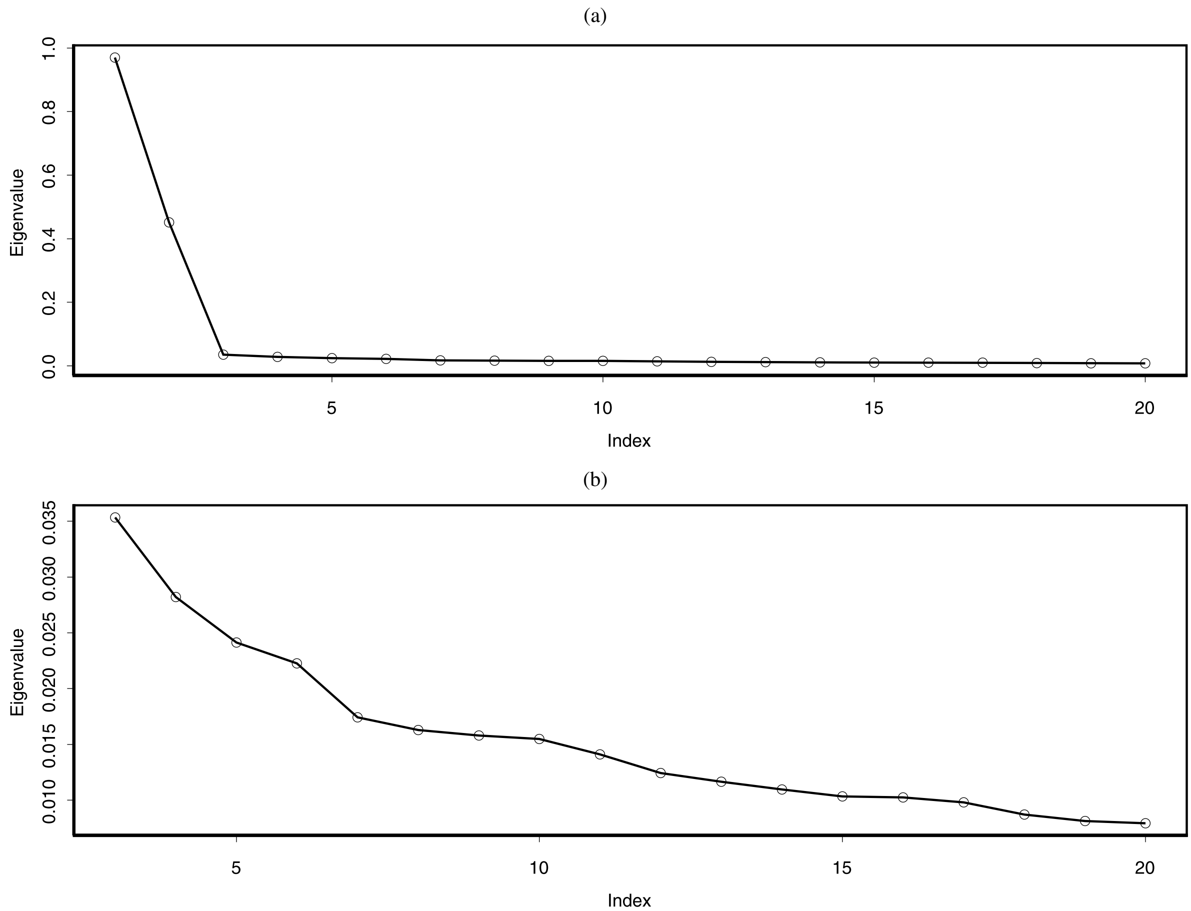

Figure 3. Plots of the 20 largest eigenvalues of $\bar{\mathcal{S}}_y$ for the dataset from Shanghai Stock Exchange. (a) The plot of all 20 largest eigenvalues. (b) The plot of the third largest to 20th largest eigenvalues.

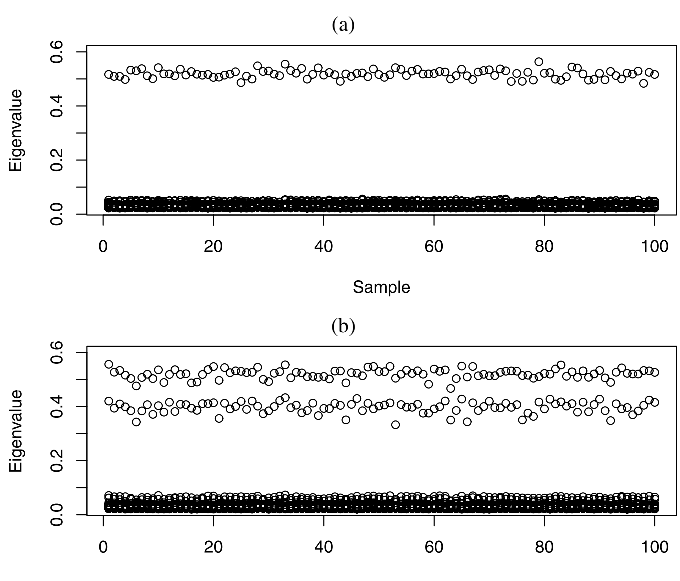

Figure 4. Plots of the 20 largest eigenvalues of $\bar{\mathcal{S}}_y$ over 100 simulated samples. The horizontal axis indicates 100 simulated samples, and the 20 largest eigenvalues of $\bar{\mathcal{S}}_y$ for each sample are plotted vertically as 20 points. (a) and (b) correspond to the cases of $r = 1$ and $r = 2$, respectively.

data. The average noise to signal ratio over 177 days is found to be 0.009 and 0.002 for the stocks traded in the Shenzhen and Shanghai markets, respectively. Noise standard deviation 0.064 used in the simulation amounts to average noise to signal ratio 0.009. To replicate the noise to signal ratio scenarios in the real data, we reduce the noise to signal ratio in the simulation study by decreasing noise standard deviation from 0.064 to 0.02, which corresponds to average noise to signal ratio from 0.009 to 0.001. Again we have discovered that the eigenvalues exhibit the resembling patterns. Moreover, we find that the smaller the noise standard deviations are, the tighter the eigenvalue clusters are.

We propose a data-dependent method to select m for ARVM estimator defined in (6) and (7) as follows. Let m be the grid number of presampling frequencies $\boldsymbol{\tau}^k$ in (5). To denote the dependence on m, we add superscript m to daily ARVM estimators given by (6) and (7) and denote them by $\tilde{\boldsymbol{\Sigma}}_y^m(\ell) = (\tilde{\Sigma}_y^m(\ell)[i_1, i_2])$ for the ℓth day, $\ell = 1, \ldots, L$. Since for each (i_1, i_2), $\tilde{\Sigma}_y^m(\ell)[i_1, i_2]$ is a daily realized covolatility between assets i_1 and i_2, we predict one day ahead daily realized covolatility by current daily realized covolatility and use predication errors as a criterion to select m. Let

$$\Psi(m) = \frac{1}{p^2 L} \sum_{i_1=1}^{p} \sum_{i_2=1}^{p} \sum_{\ell=2}^{L} \{\tilde{\Sigma}_y^m(\ell - 1)[i_1, i_2] - \tilde{\Sigma}_y^m(\ell)[i_1, i_2]\}^2.$$

The value of m is selected by minimizing $\Psi(m)$, and then used for ARVM estimator $\tilde{\mathbf{\Sigma}}_y^m(\ell)$ defined in (6) and (7).

4.3 Matrix Factor Model and VAR Model Fitting

The patterns exhibited in Figures 2 and 3 and the simulation study lead us to select $r = 1$ and $r = 2$ for the Shenzhen and Shanghai datasets, respectively. We proceed our analysis for the Shenzhen Stock Exchange data with $r = 1$. Let $\widehat{\mathbf{A}}$ be the eigenvector of $\bar{\mathcal{S}}_y$ corresponding to the largest eigenvalue. We then evaluate the factor volatility sequence $\widehat{\mathbf{\Sigma}}_f(\ell) = \widehat{\mathbf{A}}^T \widehat{\mathbf{\Sigma}}_y(\ell)\widehat{\mathbf{A}}$, $\ell = 1, \ldots, L = 177$, which is now a univariate time series. An AR(3) model, selected from PACF together with AIC and BIC, is fitted to the time series $\widehat{\mathbf{\Sigma}}_f(\ell)$. Figure 5 displays the time series plots and the ACF plots of both the original time series $\widehat{\mathbf{\Sigma}}_f(\ell)$ and the residuals resulted from the AR(3) fitting. It shows that the factor model and also the AR(3) model for factors provide reasonably good fittings to the data.

Now we move to the analysis of the Shanghai Stock Exchange data with $r = 2$. The estimator $\widehat{\mathbf{A}}$ of factor loadings $\mathbf{A}$ is taken to be the 2×630 matrix consisting of the two eigenvectors of $\bar{\mathcal{S}}_y$ corresponding to the two largest eigenvalues. Now the daily factor volatilities $\widehat{\mathbf{\Sigma}}_f(\ell) = \widehat{\mathbf{A}}^T \widehat{\mathbf{\Sigma}}_y(\ell)\widehat{\mathbf{A}}$, $\ell = 1, \ldots, L = 177$, is a series of 2×2 matrices.

Take the two diagonal elements and one off-diagonal element from $\widehat{\mathbf{\Sigma}}_f(\ell)$ to form trivariate time series $\mathrm{vech}\{\widehat{\mathbf{\Sigma}}_f(\ell)\}$, which is plotted in Figure 6. We fit $\mathrm{vech}\{\widehat{\mathbf{\Sigma}}_f(\ell)\}$ to the VAR model and use AIC and BIC criteria to select its order q.

The fitting yields a VAR model of order $q = 2$ with the estimated coefficients

$$\widehat{\alpha}_0 = \begin{pmatrix} 0.008 \\ 0.003 \\ 0.008 \end{pmatrix}, \qquad \widehat{\alpha}_1 = \begin{pmatrix} 0.016 & 0.099 & 0.162 \\ -0.232 & -0.396 & 0.822 \\ -0.407 & -0.747 & 1.218 \end{pmatrix},$$

$$\widehat{\alpha}_2 = \begin{pmatrix} 0.523 & 1.295 & -0.981 \\ 0.109 & 0.262 & -0.203 \\ 0.387 & 0.961 & -0.649 \end{pmatrix}$$

and the estimated innovation covariance matrix

$$\begin{pmatrix} 0.0045 & -0.0011 & 0.0010 \\ -0.0011 & 0.0006 & 0.0002 \\ 0.0010 & 0.0002 & 0.0007 \end{pmatrix}.$$

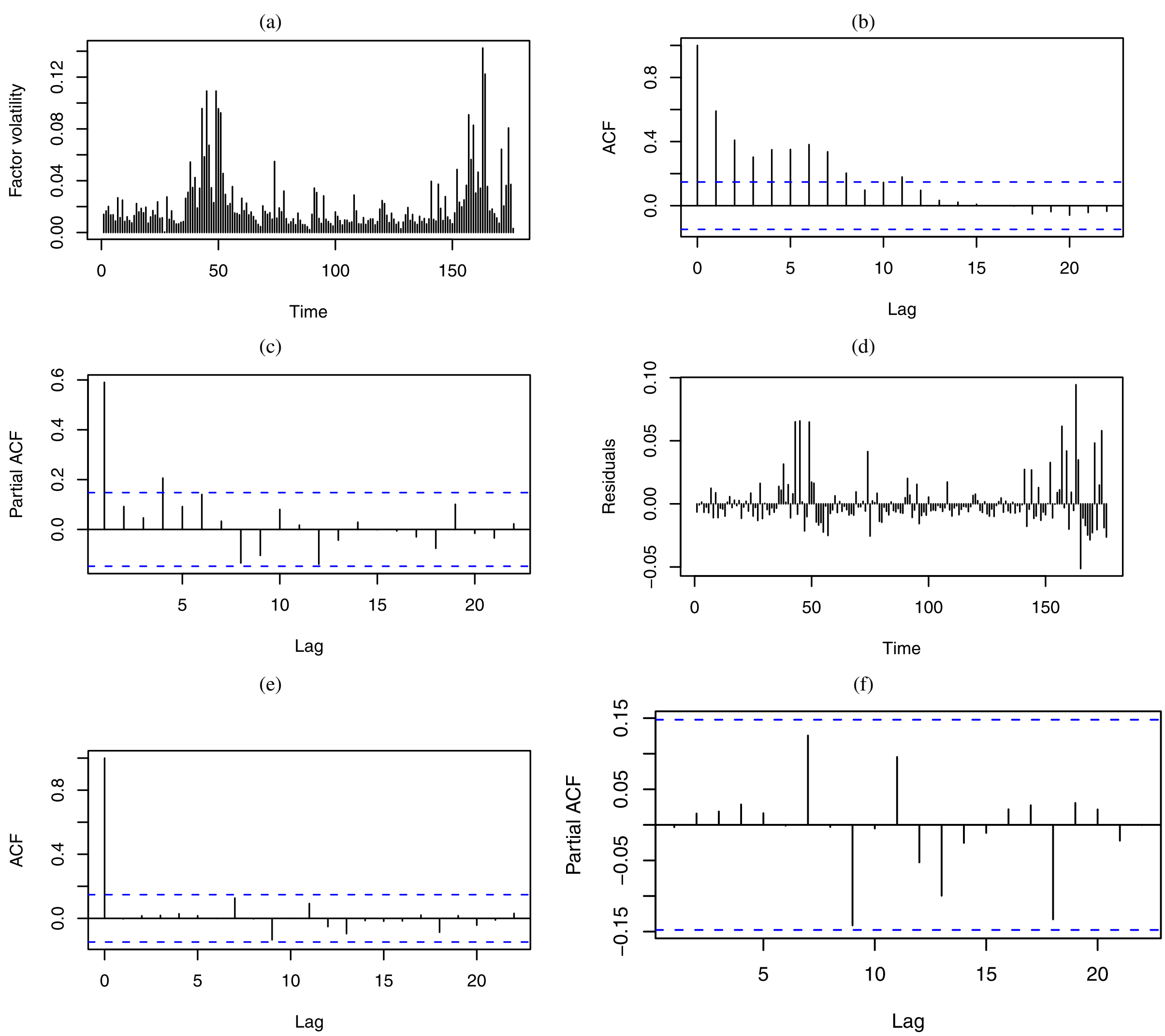

Figure 5. Fitting Shenzhen data: (a) time plot of factor volatility series, (b) ACF of factor volatility series, (c) PACF of factor volatility series, (d) time plot of the residuals from the AR(3) fitting, (d) ACF of the residuals, and (e) PACF of the residuals. The online version of this figure is in color.

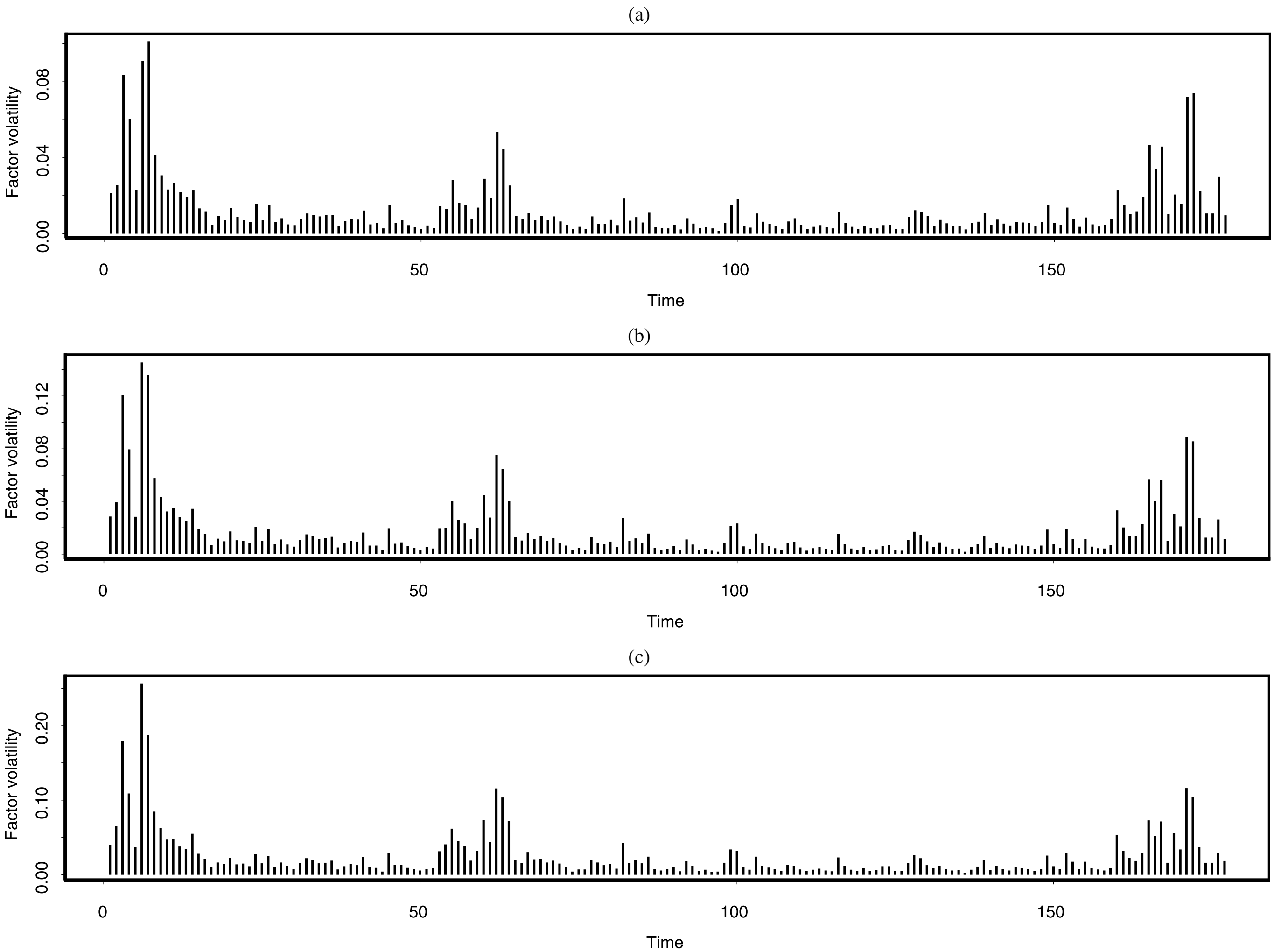

Figure 6. Time plots for vech($\widehat{\boldsymbol{\Sigma}}_f$) for the Shanghai Stock Exchange data. (a) and (b) correspond to the first and second diagonal elements of $\widehat{\boldsymbol{\Sigma}}_f$, respectively, with (c) for the off-diagonal element of $\widehat{\boldsymbol{\Sigma}}_f$.

The ACFs of vech$\{\widehat{\boldsymbol{\Sigma}}_f(\ell)\}$ plotted in Figure 7 show that the factor volatility series are highly correlated. Figure 8(a)–(c) displays the residuals resulted from above model fitting, whose ACFs are plotted in Figure 9. These plots indicate that the VAR(2) model provides adequate fit to the data.

5. CONCLUSIONS

In this article, we have proposed a novel approach to model the volatility and covolatility dynamics of daily returns for a large number of financial assets based on high-frequency intraday data. The core of the proposed method is to impose a matrix form of factor model on the sparse versions of realized volatility estimators obtained via thresholding. The fitting of the factor model boils down to an eigen-analysis for a nonnegative definite matrix, and therefore is feasible with an ordinary PC when the number of assets is in the order of a few thousands. The asymptotic theory is developed in the manner that the number of assets, the numbers of intraday observations and the number of days concerned go to infinity all together. Numerical illustration with intraday prices from both Shenzhen and Shanghai markets indicates that the factor modeling strategy works effectively as the daily volatility dynamics of all the assets in those two markets was driven by one (for Shenzhen) or two (for Shanghai) common factors.

As far as we are aware, this work represents the first attempt to use high-frequency data to model ultra-high dimensional volatility matrices and combine high-frequency volatility matrix estimation with low-frequency volatility dynamic models. While the approach yields new volatility estimation and prediction procedures that are better than methods only based on either high-frequency volatility estimation or low-frequency volatility dynamic modeling, we leave some open issues as well as a number of important future research topics. For example, volatility factors are important both statistically and economically, it is desirable to have data-driven methods to select the number of significant factors for fitting the VAR model. The ARVM estimator is used to estimate daily volatility matrices and perform eigen-analysis in Sections 2.2 and 2.3, it is very interesting and challenging to investigate the performance of the methodology when other volatility matrix estimators instead of the ARVM estimator are employed. Large volatility matrix prediction is another important research topic. For example, the fitted matrix factor and VAR(2) models obtained

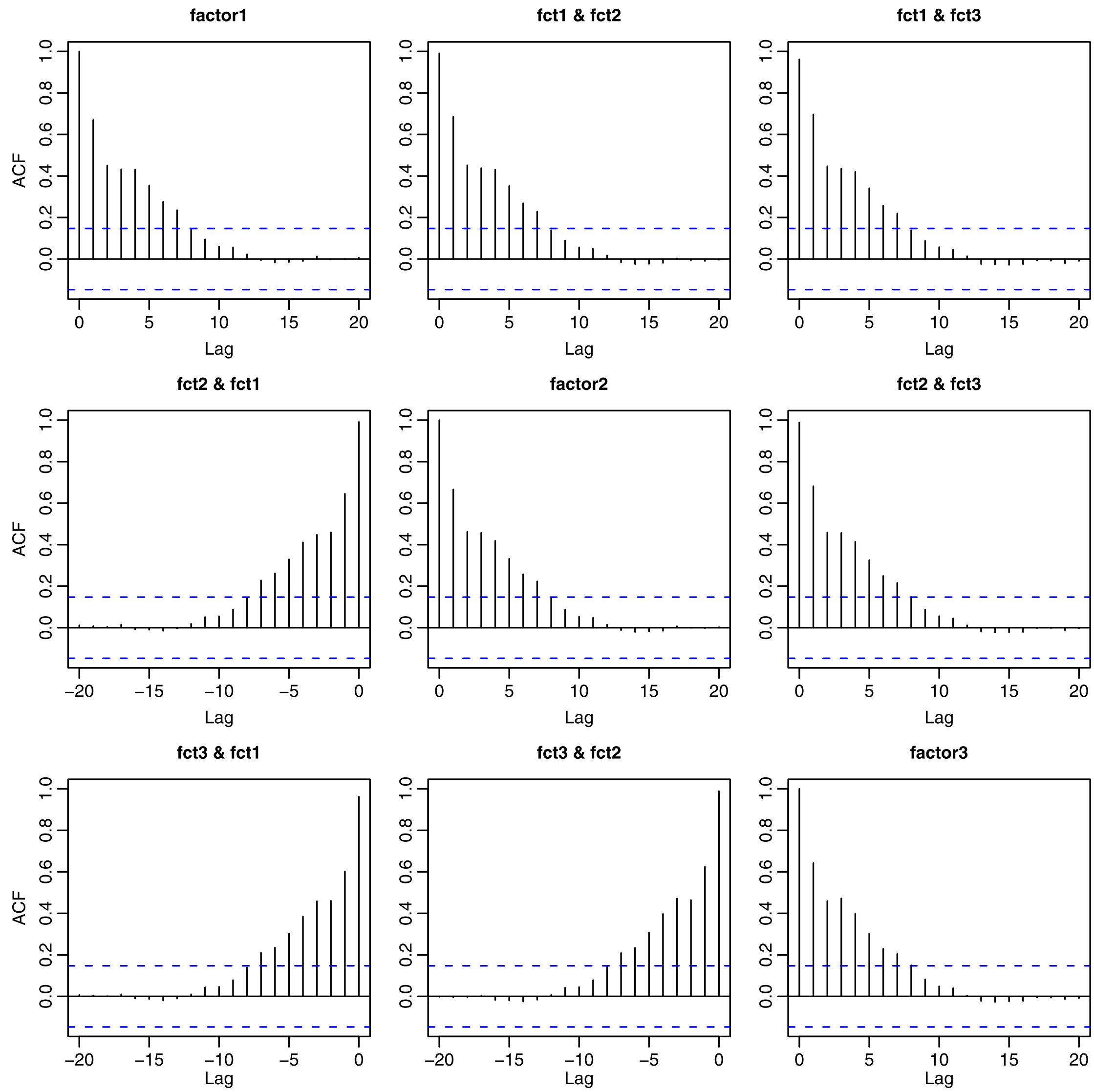

Figure 7. ACF plots of the corresponding factor volatility $\mathrm{vech}(\widehat{\boldsymbol{\Sigma}}_f)$ displayed in Figure 6 for the dataset from Shanghai Stock Exchange. The three plots on diagonal correspond to the ACFs of three factor volatility components with off-diagonal plots for their cross ACFs. The online version of this figure is in color.

from Shanghai market data can be used to forecast future integrated volatility matrix by first predicting h-step ahead factor volatility $\boldsymbol{\Sigma}_f(L+h)$ from the derived VAR(2) model and then using matrix factor model (10) to evaluate h-step ahead forecast of integrated volatility matrix $\boldsymbol{\Sigma}_x(L+h)$. However, for the prediction of large volatility matrices, we need to properly gauge the predict error and investigate the impact of matrix size on the prediction.

APPENDIX: PROOFS OF THEOREMS

Besides matrix norm, we need other two ℓ_d norms. Given a p-dimensional vector $\mathbf{x} = (x_1, \ldots, x_p)^T$ and a p by p matrix $\mathbf{U} = (U_{ij})$, define their ℓ_d-norms as follows:

$$\|\mathbf{x}\|_d = \left(\sum_{i=1}^{p} |x_i|^d\right)^{1/d},$$

$$\|\mathbf{U}\|_d = \sup\{\|\mathbf{U}\mathbf{x}\|_d, \|\mathbf{x}\|_d = 1\}, \qquad d = 1, 2, \infty.$$

Note the facts that $\|\mathbf{U}\|_2$ is equal to the square root of the largest eigenvalue of $\mathbf{U}^T\mathbf{U}$,

$$\|\mathbf{U}\|_1 = \max_{1\le j\le p} \sum_{i=1}^{p} |U_{ij}|, \qquad \|\mathbf{U}\|_\infty = \max_{1\le i\le p} \sum_{j=1}^{p} |U_{ij}|,$$

and

$$\|\mathbf{U}\|_2^2 \le \|\mathbf{U}\|_1 \|\mathbf{U}\|_\infty.$$

For symmetric $\mathbf{U}$, $\|\mathbf{U}\|_2$ is equal to its largest absolute eigenvalue, and $\|\mathbf{U}\|_2 \le \|\mathbf{U}\|_1 = \|\mathbf{U}\|_\infty$. Denote by C generic constant whose value may change from appearance to appearance.

Before proving theorems we need to establish six lemmas. Lemmas A.1 and A.2 show that Condition (A2) gives an order for $\|\boldsymbol{\Sigma}_x(\ell)\|_2$ while Condition (A1) together with (A2) guarantee sparsity for all $\boldsymbol{\Sigma}_x(\ell)$.

Lemma A.1. Assumption (A2) implies that the maximum eigenvalue of $\boldsymbol{\Sigma}_x(\ell)$ are bounded uniformly over $\ell = 1, \ldots, L$, that is,

$$\max_{1\le\ell\le L} \|\boldsymbol{\Sigma}_x(\ell)\|_2 = O_P(\log L).$$

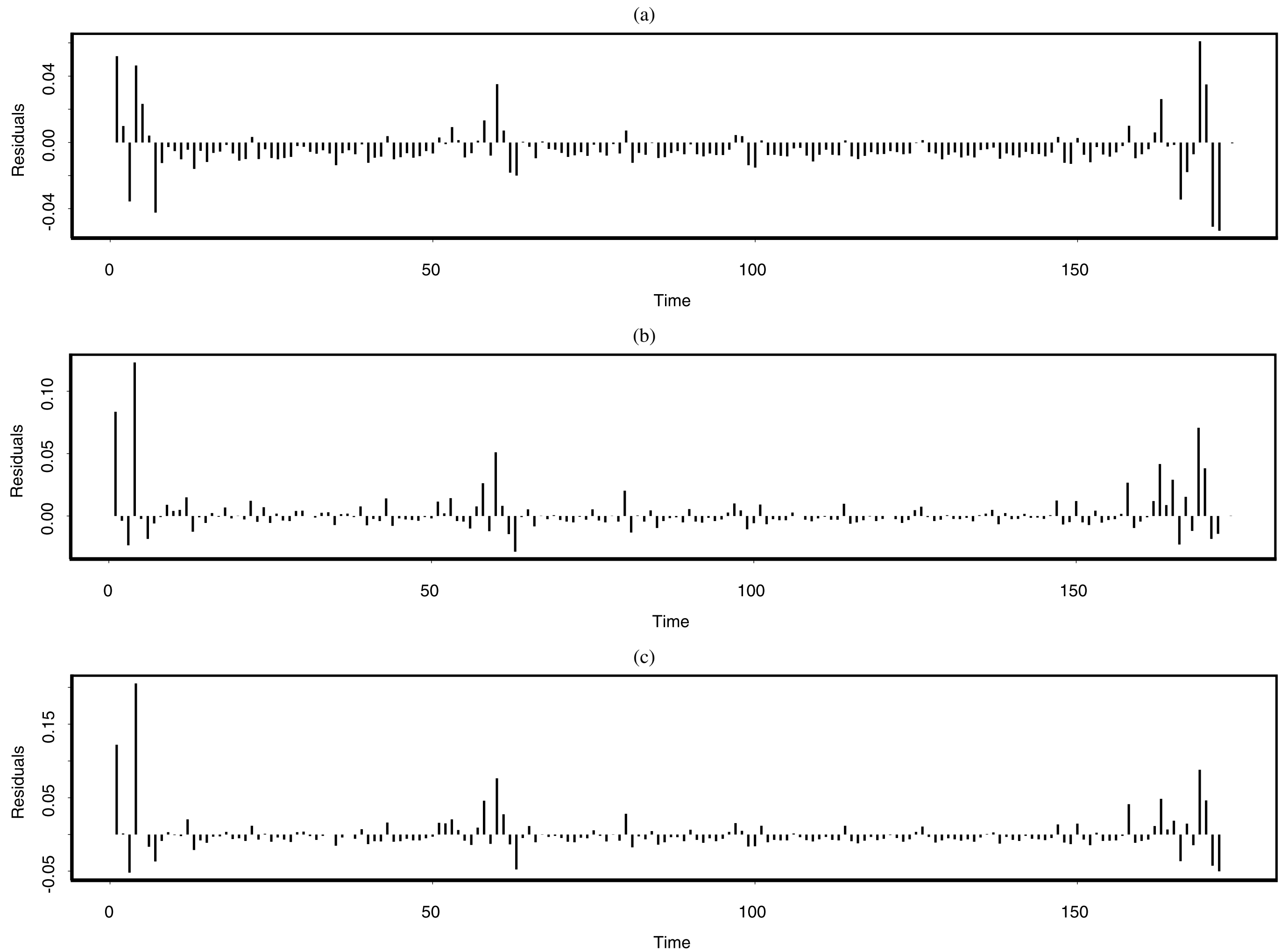

Figure 8. Time plots of the residuals resulted from a VAR(2) fitting to vech($\widehat{\boldsymbol{\Sigma}}_f$) for the Shanghai Stock Exchange data. (a) and (b) correspond to the first and second diagonal elements of $\widehat{\boldsymbol{\Sigma}}_f$, respectively, and (c) to the off-diagonal element of $\widehat{\boldsymbol{\Sigma}}_f$.

Proof. From factor model (10) and submultiplicative property of norm $\|\cdot\|_2$ (i.e., $\|\mathbf{U}\mathbf{V}\|_2 \le \|\mathbf{U}\|_2\|\mathbf{V}\|_2$ for matrices $\mathbf{U}$ and $\mathbf{V}$), we have

$$\|\boldsymbol{\Sigma}_x(\ell)\|_2 \le \|\mathbf{A}\boldsymbol{\Sigma}_f(\ell)\mathbf{A}^T + \boldsymbol{\Sigma}_0\|_2 \le \|\mathbf{A}\|_2\|\boldsymbol{\Sigma}_f(\ell)\|_2\|\mathbf{A}^T\|_2 + \|\boldsymbol{\Sigma}_0\|_2$$
$$\le r^2 \sum_{j=1}^{r} \Sigma_f(\ell)[j,j] + \|\boldsymbol{\Sigma}_0\|_2,$$

where we use the facts that since $\|\mathbf{A}^T\|_2, \|\mathbf{A}\|_2 \le \operatorname{trace}(\mathbf{A}\mathbf{A}^T) = \operatorname{trace}(\mathbf{A}^T\mathbf{A}) = r$, and $\|\boldsymbol{\Sigma}_f(\ell)\|_2 \le \operatorname{trace}(\boldsymbol{\Sigma}_f(\ell)) = \sum_{j=1}^{r} \Sigma_f(\ell)[j,j]$. The lemma is a direct consequence of Assumption (A2).

Lemma A.2. Assumptions (A1) and (A2) imply sparsity for $\boldsymbol{\Sigma}_x(\ell)$ uniformly over $\ell = 1, \ldots, L$, that is,

$$\sum_{j=1}^{p} |\Sigma_x(\ell)[i,j]|^\delta \le M\pi(p,L), \qquad i = 1,\ldots,p, \ell = 1,\ldots,L, \quad \text{(A.1)}$$

where M is a positive random variable, $\pi(p,L) = \pi(p)\log^\delta L$, and δ and $\pi(p)$ are given as in Assumption (A1).

Proof. First we give an inequality that for any $y_1, \ldots, y_m$,

$$\left(\sum_{j=1}^{m} |y_j|\right)^\delta \le \sum_{j=1}^{m} |y_j|^\delta. \qquad \text{(A.2)}$$

Take $w_j = |y_j| / \sum_{j=1}^{m} |y_j|$. Then $\sum_{j=1}^{m} w_j = 1$, $0 \le w_j \le 1$, and $w_j^\delta \ge w_j$. The inequality is proved as follows:

$$\sum_{j=1}^{m} w_j^\delta \ge \sum_{j=1}^{m} w_j = 1.$$

Inequality (A.2) indicates that the sum of two sparse matrices are also sparse. Thus with Condition (A1) and (10) it is enough to show that $\mathbf{A}\boldsymbol{\Sigma}_f(\ell)\mathbf{A}^T$ is sparse for $\ell = 1, \ldots, L$.

Let $\mathbf{A} = (a_{ij})$, $\boldsymbol{\Sigma}_f(\ell) = (\Sigma_f(\ell)[i,j])$, $\mathbf{U} = \mathbf{A}\boldsymbol{\Sigma}_f(\ell)\mathbf{A}^T = (u_{ij})$, and $G = \max\{|\Sigma_f(\ell)[i,j]|, \ell = 1,\ldots,L, i,j = 1,\ldots,r\}$. Since $\boldsymbol{\Sigma}_f(\ell)$ are positive definite, (A2) implies that $G = O_P(\log L)$. Hence,

$$\begin{aligned} |u_{ij}|^\delta &= \left|\sum_{h=1}^{r}\sum_{k=1}^{r} a_{ih}\Sigma_f(\ell)[h,k]a_{jk}\right|^\delta \\ &\le \sum_{h=1}^{r}\sum_{k=1}^{r} |a_{ih}\Sigma_f(\ell)[h,k]a_{jk}|^\delta \\ &\le G^\delta \sum_{h=1}^{r}\sum_{k=1}^{r} |a_{ih}a_{jk}|^\delta, \end{aligned} \qquad \text{(A.3)}$$

$$\sum_{j=1}^{p} |u_{ij}|^\delta \le G^\delta \sum_{h=1}^{r}\sum_{k=1}^{r} |a_{ih}|^\delta \sum_{j=1}^{p} |a_{jk}|^\delta \le r^2 C G^\delta \pi(p),$$

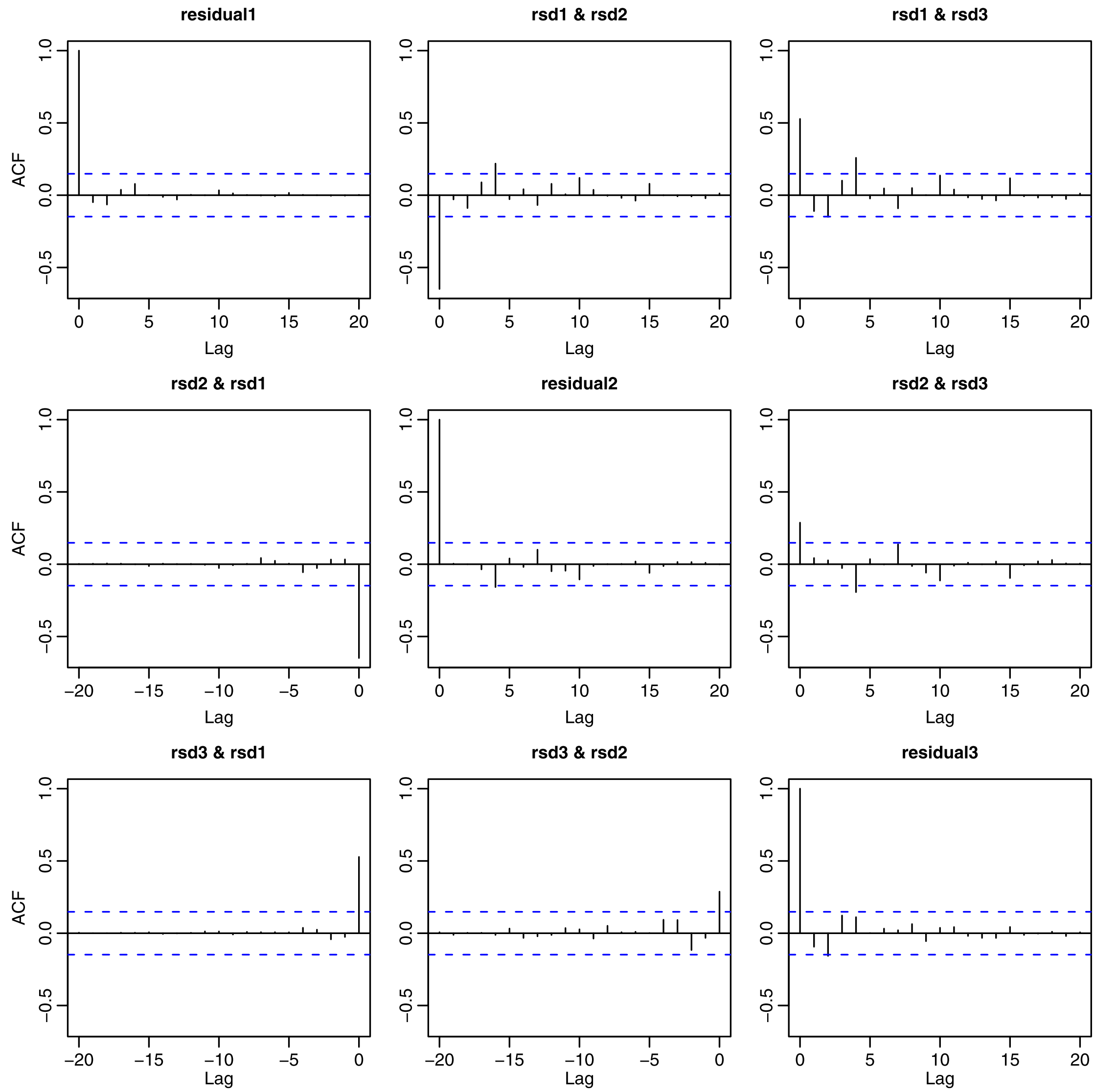

Figure 9. ACF plots of the corresponding three residual components in Figure 8 for the dataset from Shanghai Stock Exchange. The three plots on diagonal correspond to the ACFs of three residual components with off-diagonal plots for their cross ACFs. The online version of this figure is in color.

where the last inequality is from the facts that the elements of $\mathbf{A}$ are bounded by 1 and the column vectors of A obey (18). As $G = O_P(\log L)$, the bound $r^2 C G^\delta \pi(p)$ on the right-hand side of (A.3) can be expressed as $M\pi(p, L)$.

The next lemma derives the summation results under the established sparsity in Lemma A.2.

Lemma A.3. The sparsity established in Lemma A.2 for all $\boldsymbol{\Sigma}_x(\ell)$ infers that for any fixed $a > 0$,

$$\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_x(\ell)[i,j]|1\big(|\Sigma_x(\ell)[i,j]|\le a\varpi\big) = O_P(\pi(p,L)\varpi^{1-\delta}), \tag{A.4}$$

$$\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}1\big(|\Sigma_x(\ell)[i,j]|\ge a\varpi\big) = O_P(\pi(p,L)\varpi^{-\delta}). \tag{A.5}$$

Proof. With simple algebraic manipulations we obtain

$$\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_x(\ell)[i,j]|1\big(|\Sigma_x(\ell)[i,j]|\le a\varpi\big)$$

$$\le (a\varpi)^{1-\delta}\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_x(\ell)[i,j]|^\delta 1\big(|\Sigma_x(\ell)[i,j]|\le a\varpi\big)$$

$$\le (a\varpi)^{1-\delta}\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_x(\ell)[i,j]|^\delta \le (a\varpi)^{1-\delta}M\pi(p,L)$$

$$= O_P(\pi(p,L)\varpi^{1-\delta}),$$

which proves (A.4). Equation (A.5) is proved as follows:

$$\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}1\big(|\Sigma_x(\ell)[i,j]|\ge a\varpi\big)$$

$$\le \max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}\left(\frac{|\Sigma_x(\ell)[i,j]|}{a\varpi}\right)^\delta 1\big(|\Sigma_x(\ell)[i,j]|\ge a\varpi\big)$$

$$\leq (a\varpi)^{-\delta} \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} |\Sigma_x(\ell)[i,j]|^{\delta}$$

$$\leq (a\varpi)^{-\delta} M\pi(p,L) = O_P(\pi(p,L)\varpi^{-\delta}).$$

The next two lemmas are results about ARVM estimator $\tilde{\mathbf{\Sigma}}_y(\ell)$ that we need later to establish a convergence rate for TARVM estimator $\widehat{\mathbf{\Sigma}}_y(\ell)$.

Lemma A.4. Under models (1)–(2) and Conditions (A3)–(A4) we have for all $1 \leq i,j \leq p$ and $1 \leq \ell \leq L$,

$$E\big(|\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]|^{\beta}\big) \leq C e_n^{\beta}, \tag{A.6}$$

where C is a generic constant free of n, p, and L, and the convergence rate e_n is specified as $e_n \sim n^{-1/6}$ for the noise case and $e_n \sim n^{-1/3}$ for the noiseless case [i.e., $\varepsilon_i(t_{ij}) = 0$ in (2)].

Proof. The lemma is a consequence of applying Theorem 1 in Wang and Zou (2010) to the current set-up.

Lemma A.5. Under Conditions (A1)–(A4), we have

$$\max_{1\leq\ell\leq L} \max_{1\leq i,j\leq p} |\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]|$$

$$= O_P\big(e_n(p^2L)^{1/\beta}\big) = o_P(\varpi), \tag{A.7}$$

$$P\left(\max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\{|\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]| \geq \varpi/2\} > 0\right)$$

$$= o(1), \tag{A.8}$$

$$\max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\big(|\tilde{\Sigma}_y(\ell)[i,j]| \geq \varpi, |\Sigma_x(\ell)[i,j]| < \varpi\big)$$

$$= O_P(\pi(p)\varpi^{-\delta}), \tag{A.9}$$

where ϖ is as in Theorem 1.

Proof. Taking $d = d_1 e_n (p^2L)^{1/\beta}$ and applying Markov inequality and (A.6), we have

$$P\Big(\max_{1\leq\ell\leq L} \max_{1\leq i,j\leq p} |\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]| > d\Big)$$

$$\leq \sum_{\ell=1}^{L} \sum_{i,j=1}^{p} P\big(|\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]| > d\big)$$

$$\leq \frac{Cp^2Le_n^{\beta}}{d^{\beta}} = \frac{C}{d_1^{\beta}} \to 0,$$

as $p,n,L \to \infty$ and then $d_1 \to \infty$. This proves (A.7), using which we can obtain

$$P\left(\max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\{|\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]| \geq \varpi/2\} > 0\right)$$

$$\leq P\Big(\max_{1\leq\ell\leq L} \max_{1\leq i,j\leq p} |\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]| \geq \varpi/2\Big)$$

$$\leq \frac{2^{\beta}p^2LCe_n^{\beta}}{\varpi^{\beta}} = \frac{2^{\beta}C}{\log^{\beta} L} \to 0,$$

as $n,p,L \to 0$, which proves (A.8). Then we apply (A.5) and (A.8) to show (A.9) as follows.

$$\max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\big(|\tilde{\Sigma}_y(\ell)[i,j]| \geq \varpi, |\Sigma_x(\ell)[i,j]| < \varpi\big)$$

$$\leq \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\big(|\tilde{\Sigma}_y(\ell)[i,j]| \geq \varpi, |\Sigma_x(\ell)[i,j]| \leq \varpi/2\big)$$

$$+ \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\big(|\tilde{\Sigma}_y(\ell)[i,j]| \geq \varpi,$$

$$\varpi/2 < |\Sigma_x(\ell)[i,j]| < \varpi\big)$$

$$\leq \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\big(|\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]| \geq \varpi/2\big)$$

$$+ \max_{1\leq l\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} 1\big(|\Sigma_x(\ell)[i,j]| > \varpi/2\big)$$

$$\leq o_P(1) + 2^{\delta} M\pi(p,L)\varpi^{-\delta} = O_P(\pi(p,L)\varpi^{-\delta}).$$

The next lemma provides the convergence rate for TARVM estimator $\widehat{\mathbf{\Sigma}}_y(\ell)$ under matrix norm uniformly over all ℓ.

Lemma A.6. Under Conditions (A1)–(A4) we have

$$\max_{1\leq\ell\leq L} \|\widehat{\mathbf{\Sigma}}_y(\ell) - \mathbf{\Sigma}_x(\ell)\|_2 = O_P(\pi(p,L)\varpi^{1-\delta})$$

$$= O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta} \log L\big),$$

where e_n and ϖ are as in Theorem 1.

Proof. Using the relationship between ℓ_2 and ℓ_∞ norms and triangle inequality, we have

$$\max_{1\leq\ell\leq L} \|\widehat{\mathbf{\Sigma}}_y(\ell) - \mathbf{\Sigma}_x(\ell)\|_2 \leq \max_{1\leq\ell\leq L} \|\widehat{\mathbf{\Sigma}}_y(\ell) - \mathbf{\Sigma}_x(\ell)\|_\infty$$

$$\leq \underbrace{\max_{1\leq\ell\leq L} \big\|\widehat{\mathbf{\Sigma}}_y(\ell) - \mathcal{T}_{\varpi}[\mathbf{\Sigma}_x(\ell)]\big\|_\infty}_{I}$$

$$+ \underbrace{\max_{1\leq\ell\leq L} \big\|\mathcal{T}_{\varpi}[\mathbf{\Sigma}_x(\ell)] - \mathbf{\Sigma}_x(\ell)\big\|_\infty}_{II}.$$

Lemma A.3 implies

$$II = \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} |\Sigma_x(\ell)[i,j]| 1\big(|\Sigma_x(\ell)[i,j]| \leq \varpi\big)$$

$$= O_P(\pi(p,L)\varpi^{1-\delta}).$$

This lemma is proved by showing that I is also of order $\pi(p,L)\varpi^{1-\delta}$ in probability. Indeed, we have

$$I \leq \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} |\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]|$$

$$\times 1\big(|\tilde{\Sigma}_y(\ell)[i,j]| \geq \varpi, |\Sigma_x(\ell)[i,j]| \geq \varpi\big)$$

$$+ \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} |\tilde{\Sigma}_y(\ell)[i,j]|$$

$$\times 1\big(|\tilde{\Sigma}_y(\ell)[i,j]| \geq \varpi, |\Sigma_x(\ell)[i,j]| < \varpi\big)$$

$$+ \max_{1\leq\ell\leq L} \max_{1\leq i\leq p} \sum_{j=1}^{p} |\Sigma_x(\ell)[i,j]|$$

$$\times 1\big(|\tilde{\Sigma}_y(\ell)[i,j]| < \varpi, |\Sigma_x(\ell)[i,j]| \geq \varpi\big)$$

$$\leq \max_{1\leq\ell\leq L} \max_{1\leq i,j\leq p} |\tilde{\Sigma}_y(\ell)[i,j] - \Sigma_x(\ell)[i,j]|$$

$$
\begin{aligned}
&\times \max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p} 1\big(|\Sigma_x(\ell)[i,j]|\ge\varpi\big)\\
&+\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_x(\ell)[i,j]|1\big(|\Sigma_x(\ell)[i,j]|<\varpi\big)\\
&+\max_{1\le\ell\le L}\max_{1\le i,j\le p}|\tilde{\Sigma}_y(\ell)[i,j]-\Sigma_x(\ell)[i,j]|\\
&\times \max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p} 1\big(|\tilde{\Sigma}_y(\ell)[i,j]|\ge\varpi, |\Sigma_x(\ell)[i,j]|<\varpi\big)\\
&+\varpi\max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p} 1\big(|\Sigma_x(\ell)[i,j]|\ge\varpi\big)\\
&+\max_{1\le\ell\le L}\max_{1\le i,j\le p}|\tilde{\Sigma}_y(\ell)[i,j]-\Sigma_x(\ell)[i,j]|\\
&\times \max_{1\le\ell\le L}\max_{1\le i\le p}\sum_{j=1}^{p} 1\big(|\Sigma_x(\ell)[i,j]|\ge\varpi\big)\\
&= o_P(\varpi)O_p(\pi(p,L)\varpi^{-\delta})+O_p(\pi(p,L)\varpi^{1-\delta})\\
&\quad + o_P(\varpi)O_p(\pi(p,L)\varpi^{-\delta})+\varpi O_p(\pi(p,L)\varpi^{-\delta})\\
&= O_p(\pi(p,L)\varpi^{1-\delta}) = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log L\big),
\end{aligned}
$$

where the orders in the second to last equality are due to (A.4), (A.5), (A.7), and (A.9).

Remark A.1. As we have discussed in Remark 2 after Theorems 1 and 2 in Section 3, the convergence rate in Lemma A.6 indicates that for reasonably large β in moment Assumption (A3), $\widehat{\boldsymbol{\Sigma}}_y(\ell)$ provide consistent estimators of $\boldsymbol{\Sigma}_x(\ell)$ under matrix norm for large p and n. As a consequence, $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ defined in (15) are consistent estimators of $\boldsymbol{\Sigma}_f(\ell)$ under the matrix norm and in particular, with probability tending to one, $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ are semipositive definite. For finite samples, to ensure the semipositive definiteness of $\widehat{\Sigma}_y$ we may simply replace the negative eigenvalues of $\widehat{\Sigma}_y$ by zero, and hence $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ are semipositive definite. Thus, we may build a VAR model for $\boldsymbol{\Sigma}_f^{1/2}(\ell)$ instead of $\boldsymbol{\Sigma}_f(\ell)$ and fit the model to $\widehat{\boldsymbol{\Sigma}}_f^{1/2}(\ell)$.

Proof of Theorem 1

Due to the triangle inequality and submultiplicative property of norm $\|\cdot\|_2$, we have

$$
\begin{aligned}
\|\bar{\mathcal{S}}_y-\bar{\mathcal{S}}_x\|_2 &= \left\|\frac{1}{L}\sum_{\ell=1}^{L}\{\widehat{\boldsymbol{\Sigma}}_y(\ell)-\bar{\boldsymbol{\Sigma}}_y\}^2-\frac{1}{L}\sum_{\ell=1}^{L}\{\boldsymbol{\Sigma}_x(\ell)-\bar{\boldsymbol{\Sigma}}_x\}^2\right\|_2\\
&= \left\|\frac{1}{L}\sum_{\ell=1}^{L}[\widehat{\boldsymbol{\Sigma}}_y(\ell)]^2-\bar{\boldsymbol{\Sigma}}_y^2-\frac{1}{L}\sum_{\ell=1}^{L}\boldsymbol{\Sigma}_x^2(\ell)+\bar{\boldsymbol{\Sigma}}_x^2\right\|_2\\
&\le \left\|\frac{1}{L}\sum_{\ell=1}^{L}[\widehat{\boldsymbol{\Sigma}}_y(\ell)]^2-\frac{1}{L}\sum_{\ell=1}^{L}\boldsymbol{\Sigma}_x^2(\ell)\right\|_2+\|\bar{\boldsymbol{\Sigma}}_y^2-\bar{\boldsymbol{\Sigma}}_x^2\|_2\\
&\le \frac{1}{L}\sum_{\ell=1}^{L}\|\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)\|_2\cdot\big\{\|\widehat{\boldsymbol{\Sigma}}_y(\ell)\|_2+\|\boldsymbol{\Sigma}_x(\ell)\|_2\big\}\\
&\quad+\left(\frac{1}{L}\sum_{\ell=1}^{L}\|\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)\|_2\right)\\
&\quad\times\left(\frac{1}{L}\sum_{\ell=1}^{L}\big\{\|\widehat{\boldsymbol{\Sigma}}_y(\ell)\|_2+\|\boldsymbol{\Sigma}_x(\ell)\|_2\big\}\right),\\
&\le 2\max_{1\le\ell\le L}|\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)|_2\\
&\quad\times\Big(\max_{1\le\ell\le L}\|\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)\|_2+2\max_{1\le l\le L}\|\boldsymbol{\Sigma}_x(\ell)\|_2\Big),
\end{aligned}
$$

which can be easily shown to have order

$$
\begin{aligned}
\pi(p,L)\varpi^{1-\delta}\log L &= \pi(p)\varpi^{1-\delta}\log^{1+\delta}L\\
&\sim \pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L
\end{aligned}
$$

in probability from an application of Lemmas A.1, A.2, and A.6. The proof is completed.

Proof of Theorem 2

First we show

$$
\max_{1\le j\le r}|\widehat{\lambda}_j-\lambda_j| = O_P\big(\pi(p)[e_n(p^2L)^{1/\beta}]^{1-\delta}\log^2 L\big), \quad (A.10)
$$

$$
\max_{1\le j\le r}\|\widehat{\mathbf{a}}_j-\mathbf{a}_j\|_2 = O_P\big(\pi(p)[e_n(p^2L)^{1/\beta}]^{1-\delta}\log^2 L\big). \quad (A.11)
$$

Since $\|\cdot\|_2$ is equal to the largest absolute eigenvalue, and the top r eigenvalues of $\bar{\mathcal{S}}_x$ are separated by a constant c, thus

$$
\max_{1\le j\le r}|\widehat{\lambda}_j-\lambda_j|\le\|\bar{\mathcal{S}}_y-\bar{\mathcal{S}}_x\|_2,
$$

and (A.10) is a consequence of Theorem 1. The second result (A.11) follows directly from Theorem 1 and the same argument in the proof of theorem 5 in Bickel and Levina (2008a) (or theorem 6.1 of Kato 1966). Now we will use (A.10) and (A.11) to prove the two results in Theorem 2. From (A.11) we have for diagonal entry j of $\mathbf{A}^T\widehat{\mathbf{A}}$,

$$
\mathbf{a}_j^T\widehat{\mathbf{a}}_j = 1-\|\widehat{\mathbf{a}}_j-\mathbf{a}_j\|^2/2 = 1+O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big),
$$

and for off-diagonal entry (k,j) $(k\ne j)$,

$$
\begin{aligned}
|\mathbf{a}_k^T\widehat{\mathbf{a}}_j| &= |\mathbf{a}_k^T(\widehat{\mathbf{a}}_j-\mathbf{a}_j)|\le\|\mathbf{a}_k^T\|_2\|\widehat{\mathbf{a}}_j-\mathbf{a}_j\|_2\\
&= \|\widehat{\mathbf{a}}_j-\mathbf{a}_j\|_2 = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big).
\end{aligned}
$$

To prove the second result in Theorem 2, we use factor model (10) and estimator $\widehat{\boldsymbol{\Sigma}}_f$ in (15) to obtain

$$
\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}_f(\ell)-\boldsymbol{\Sigma}_f(\ell)-\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}\\
&= \widehat{\mathbf{A}}^T\{\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)\}\widehat{\mathbf{A}}+\widehat{\mathbf{A}}^T\boldsymbol{\Sigma}_x(\ell)\widehat{\mathbf{A}}-\boldsymbol{\Sigma}_f(\ell)-\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}\\
&= \widehat{\mathbf{A}}^T[\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)]\widehat{\mathbf{A}}+\{(\mathbf{A}^T\widehat{\mathbf{A}})^T\boldsymbol{\Sigma}_f(\ell)\mathbf{A}^T\widehat{\mathbf{A}}-\boldsymbol{\Sigma}_f(\ell)\}\\
&\quad+\{\widehat{\mathbf{A}}^T\boldsymbol{\Sigma}_0\widehat{\mathbf{A}}-\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}\}. \qquad (A.12)
\end{aligned}
$$

For the first term on the right-hand side of (A.12), since

$$
\big\|\widehat{\mathbf{A}}^T[\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)]\widehat{\mathbf{A}}\big\|_2\le\|\widehat{\mathbf{A}}^T\|_2\|\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)\|_2\|\widehat{\mathbf{A}}\|_2,
$$

and the columns of $\widehat{\mathbf{A}}$ are orthonormal vectors, we have

$$
\|\widehat{\mathbf{A}}^T\|_2^2, \|\widehat{\mathbf{A}}\|_2^2\le \mathrm{trace}(\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T) = \mathrm{trace}(\widehat{\mathbf{A}}^T\widehat{\mathbf{A}}) = r.
$$

From Theorem 1, we conclude

$$
\begin{aligned}
\big\|\widehat{\mathbf{A}}^T[\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)]\widehat{\mathbf{A}}\big\|_2 &\le \|\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)\|_2\\
&= O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big).
\end{aligned}
$$

As $\widehat{\mathbf{A}}^T[\widehat{\boldsymbol{\Sigma}}_y(\ell)-\boldsymbol{\Sigma}_x(\ell)]\widehat{\mathbf{A}}$ is r by r matrix, matrix norm convergence implies convergence in element, so the first term is proved to be of a desired order. Note $\boldsymbol{\Sigma}_f(\ell)$ are r by r matrices, from Condition (A2) we easily conclude that the second term on the right-hand side of (A.12) is of the order $\mathbf{A}^T\widehat{\mathbf{A}} - \mathbf{I}_r$, which has the requested order. For the third term on the right-hand side of (A.12) we have

$$\begin{aligned}
&\|\widehat{\mathbf{A}}^T\boldsymbol{\Sigma}_0\widehat{\mathbf{A}} - \mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}\|_2 \\
&\quad \le \|(\widehat{\mathbf{A}} - \mathbf{A})^T\boldsymbol{\Sigma}_0\widehat{\mathbf{A}} + \mathbf{A}^T\boldsymbol{\Sigma}_0(\widehat{\mathbf{A}} - \mathbf{A})\|_2 \\
&\quad \le \|(\widehat{\mathbf{A}} - \mathbf{A})^T\boldsymbol{\Sigma}_0\widehat{\mathbf{A}}\|_2 + \|\mathbf{A}^T\boldsymbol{\Sigma}_0(\widehat{\mathbf{A}} - \mathbf{A})\|_2 \\
&\quad \le \|(\widehat{\mathbf{A}} - \mathbf{A})^T\|_2\|\boldsymbol{\Sigma}_0\|_2\|\widehat{\mathbf{A}}\|_2 + \|\mathbf{A}^T\|_2\|\boldsymbol{\Sigma}_0\|_2\|(\widehat{\mathbf{A}} - \mathbf{A})\|_2 \\
&\quad = \|\widehat{\mathbf{A}} - \mathbf{A}\|_2\|\boldsymbol{\Sigma}_0\|_2[\|\widehat{\mathbf{A}}\|_2 + \|\mathbf{A}\|_2\|].
\end{aligned}$$

Condition (A2) guarantees that $\|\boldsymbol{\Sigma}_0\|_2$ is bounded, it has been shown that $\|\mathbf{A}\|_2 \le r$ and $\|\widehat{\mathbf{A}}\|_2 \le r$, and

$$\begin{aligned}
\|\widehat{\mathbf{A}} - \mathbf{A}\|_2^2 &\le \operatorname{trace}(\widehat{\mathbf{A}} - \mathbf{A})(\widehat{\mathbf{A}} - \mathbf{A})^T = \operatorname{trace}(\widehat{\mathbf{A}} - \mathbf{A})^T(\widehat{\mathbf{A}} - \mathbf{A}) \\
&= 2\operatorname{trace}(\mathbf{I}_r - \mathbf{A}^T\widehat{\mathbf{A}}) = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big).
\end{aligned}$$

Therefore, the third term in (A.12) is also of correct order. With all three terms on the right-hand side of (A.12) of order $\pi(p)[e_n(p^2 \times L)^{1/\beta}]^{1-\delta}\log^2 L$ in probability, we establish the second result in the theorem.

Proof of Theorem 3

As $\tilde{\boldsymbol{\alpha}}_i$ are the standard least squares estimators of $\boldsymbol{\alpha}_i$ in the VAR model (17) based on oracle data $\boldsymbol{\Sigma}_f(\ell)$, asymptotic theory for the VAR model shows that as $L \to \infty$,

$$L^{1/2}(\tilde{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_0, \ldots, \tilde{\boldsymbol{\alpha}}_q - \boldsymbol{\alpha}_q) \tag{A.13}$$

converges in distribution to a zero mean multivariate normal distribution. With

$$\widehat{\boldsymbol{\Sigma}}_f(\ell) = \widehat{\mathbf{A}}^T\widehat{\boldsymbol{\Sigma}_y}(\ell)\widehat{\mathbf{A}},$$

from Theorem 2, we have

$$\widehat{\boldsymbol{\Sigma}}_f(\ell) = \boldsymbol{\Sigma}_f(\ell) + \mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A} + O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big).$$

Since $\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}$ is a constant matrix free of ℓ, $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ obeys the same VAR model (17) for $\boldsymbol{\Sigma}_f(\ell)$ with an extra constant vech$[\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}]$ adding to $\boldsymbol{\alpha}_0$ and a negligible error term of order $\pi(p)[e_n(p^2 \times L)^{1/\beta}]^{1-\delta}\log^2 L$. Plugging $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ into the expressions of the least squares estimators of coefficients $\boldsymbol{\alpha}_i$ in the VAR model we immediately show that the least squares estimators based on $\widehat{\boldsymbol{\Sigma}}_f(\ell)$ and oracle data $\boldsymbol{\Sigma}_f(\ell)$ satisfy

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_0 - \tilde{\boldsymbol{\alpha}}_0 - \operatorname{vech}(\mathbf{A}^T\boldsymbol{\Sigma}_0\mathbf{A}) = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big), \\
&\widehat{\boldsymbol{\alpha}}_i - \tilde{\boldsymbol{\alpha}}_i = O_P\big(\pi(p)\big[e_n(p^2L)^{1/\beta}\big]^{1-\delta}\log^2 L\big), \qquad i = 1, \ldots, q.
\end{aligned}$$

The common limiting distribution stated in the theorem is a sequence of above results and (A.13).

[Received May 2010. Revised February 2011.]

REFERENCES

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2003), "Some Like It Smooth, and Some Like It Rough: Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility," manuscript, University of Pennsylvania.[1026,1028]

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), "Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536. [1030]

——— (2010), "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices With Noise and Non-Synchronous Trading," preprint. [1025,1030]

Bickel, P. J., and Levina, E. (2008a), "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36, 199–277. [1025,1027,1039]

——— (2008b), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577–2604. [1025,1027]

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010), "Pre-Averaging Estimators of the Ex-Post Covariance Matrix in Noisy Diffusion Models With Non-Synchronous Data," *Journal of Econometrics*, 159, 116–133. [1025,1030]

Corsi, F. (2009), "A Simple Long Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7, 174–196. [1026,1028]

Engle, R. F., Ng, V. K., and Rothschild, M. (1990), "Asset Pricing With a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45, 213–238. [1028]

Fan, J., and Wang, Y. (2007), "Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data," *Journal of the American Statistical Association*, 102, 1349–1362. [1030]

Fan, J., Fan, Y., and Lv, J. (2008), "High Dimensional Covariance Matrix Estimation Using a Factor Model," *Journal of Econometrics*, 147, 186–197. [1025]

Griffin, J. E., and Oomen, R. C. (2011), "Covariance Measurement in the Presence of Non-Synchronous Trading and Market Microstructure Noise," *Journal of Econometrics*, 160, 58–68. [1030]

Hansen, P. R., and Lunde, A. (2006), "Realized Variance and Market Microstructure Noise" (with discussion), *Journal of the Business and Economic Statistics*, 24, 127–218. [1031]

Hayashi, T., and Yoshida, N. (2005), "On Covariance Estimation of Non-Synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379. [1025]

Hautsch, N., Kyj, R. M., and Oomen, L. C. A. (2009), "A Blocking and Regularization Approach to High Dimensional Realized Covariance Estimation," available at *http://ideas.repec.org/p/hum/wpaper/sfb649dp2009-049.html*. [1030]

Johnstone, I. M. (2001), "On the Distribution of the Largest Eigenvalue in Principal Component Analysis," *The Annals of Statistics*, 29, 295–327. [1025]

Johnstone, I. M., and Lu, A. Y. (2009), "On Consistency and Sparsity for Principal Component Analysis in High Dimensions" (with discussions), *Journal of the American Statistical Association*, 104, 682–703. [1025,1027]

Kato, T. (1966), *Perturbation Theory for Linear Operators*, Berlin: Springer. [1039]

Pan, J., and Yao, Q. (2008), "Modelling Multiple Time Series via Common Factors," *Biometrika*, 95, 365–379. [1028]

Tao, M., Wang, Y., and Chen, X. (2011), "Fast Convergence Rates in Estimating Large Volatility Matrices Using High-Frequency Financial Data," to appear. [1030]

Wang, M., and Yao, Q. (2005), "Modelling Multivariate Volatilities: An ad hoc Method," in *Contemporary Multivariate Analysis and Design of Experiments: In Celebration of Prof. Kai-Tai Fang's 65th Birthday*, eds. J. Fan and G. Li, Singapore: World Scientific, pp. 87–97. [1028]

Wang, Y. (2002), "Asymptotic Nonequivalence of ARCH Models and Diffusions," *The Annals of Statistics*, 30, 754–783. [1025]

Wang, Y., and Zou, J. (2010), "Vast Volatility Matrix Estimation for High-Frequency Financial Data," *The Annals of Statistics*, 38, 943–978. [1025, 1027,1031,1038]

Zhang, L. (2006), "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043. [1030]

——— (2011), "Estimating Covariation: Epps Effect, Microstructure Noise," *Journal of Econometrics*, 160, 33–47. [1025,1030]