Econometric Theory, **29**, 2013, 838–856.
doi:10.1017/S0266466612000746

FAST CONVERGENCE RATES IN ESTIMATING LARGE VOLATILITY MATRICES USING HIGH-FREQUENCY FINANCIAL DATA

MINJING TAO and YAZHEN WANG
University of Wisconsin-Madison

XIAOHONG CHEN
Yale University

Financial practices often need to estimate an integrated volatility matrix of a large number of assets using noisy high-frequency data. Many existing estimators of a volatility matrix of small dimensions become inconsistent when the size of the matrix is close to or larger than the sample size. This paper introduces a new type of large volatility matrix estimator based on nonsynchronized high-frequency data, allowing for the presence of microstructure noise. When both the number of assets and the sample size go to infinity, we show that our new estimator is consistent and achieves a fast convergence rate, where the rate is optimal with respect to the sample size. A simulation study is conducted to check the finite sample performance of the proposed estimator.

1. INTRODUCTION

High-frequency financial data provide academic researchers and industry practitioners with an incredible experiment for analyzing financial markets, in particular for understanding market microstructure and estimating market volatility. With high-frequency data, researchers are able to estimate volatilities directly from the data and to better model the volatility dynamics. There is already rapidly growing literature on volatility estimation using high-frequency data. These estimation methods include Kristensen (2010) and Zhang, Mykland, and Aït-Sahalia (2005) for the single asset case and Ait-Sahalia, Fan, and Xiu (2010), Christensen, Kinnebrock, and Podolskij (2010), and Zhang (2011) for the multiple asset case.

All these existing volatility estimators perform well for a single asset or a small number of assets. However, when estimating volatility matrices of a large number of assets, Wang and Zou (2010) and Tao, Wang, Yao, and Zou (2011) show that the existing volatility matrix estimators have poor performance and in fact are inconsistent when both the number of assets and the sample sizes go to infinity.

Chen and Wang were partially supported by the NSF grants SES-0838161 and DMS-10563, respectively. Address correspondence to Yazhen Wang, Dept. of Statistics, University of Wisconsin, Medical Science Center, 1300 University Ave., Madison, WI 53706, USA; e-mail: yzwang@stat.wisc.edu.

 © Cambridge University Press 2013

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

To the best of our knowledge, Wang and Zou (2010) is the first to propose a consistent estimator for large integrated volatility matrices using high-frequency data, allowing for both the number of assets and the sample sizes to go to infinity. Tao et al. studied the dynamics of volatility matrices for a large number of assets by combining both high-frequency and low-frequency approaches. The convergence rates established for the large-matrix estimators in Wang and Zou and in Tao et al. depend on sample sizes with $1/6$-exponent, and hence are suboptimal.

In this paper we propose a new estimator of large integrated volatility matrices using high-frequency data, and we prove that when both the number of assets and we the sample sizes go to infinity, it can achieve a fast convergence rate that depends on sample size with $1/4$-exponent, which is optimal in the presence of microstructure noise. Our asymptotic theory is established under the general diffusion setup with microstructure noise in the data and realistic finite moment condition on asset prices, instead of the restrictive Gaussian or sub-Gaussian conditions imposed in the current statistics literature on large covariance matrix estimation.

The rest of the paper proceeds as follows. Section 2 describes the large dimensional price process and the data structure. Section 3 presents the new estimator. Section 4 establishes the fast convergence rate for the proposed estimator. Section 5 provides numerical results to illustrate the finite-sample performance of the estimator. All the proofs are given in Section 6.

2. THE MODEL SETUP

Let p be the number of assets in the study and denote by $X_i(t)$ the true log price at time t of the ith asset, $i = 1, \ldots, p$. Denote by $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^T$ the vector of the true log prices at time t of p assets. The common approach in finance assumes that $\mathbf{X}(t)$ follows a continuous-time diffusion model,

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

where $\boldsymbol{\mu}_t = (\mu_1(t), \ldots, \mu_p(t))^T$ is the drift, $\mathbf{B}_t = (B_{1t}, \ldots, B_{pt})^T$ is a standard p-dimensional Brownian motion, and $\boldsymbol{\sigma}_t$ is a p-by-p matrix with $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}_t^T\, \boldsymbol{\sigma}_t$ being the volatility matrix of $\mathbf{X}(t)$. The parameter of interest is the integrated volatility matrix

$$\boldsymbol{\Gamma} = \left(\Gamma_{ij}\right)_{1 \le i, j \le p} = \int_0^1 \boldsymbol{\gamma}(t)\, dt = \int_0^1 \boldsymbol{\sigma}_t^T\, \boldsymbol{\sigma}_t\, dt.$$

Instead of observing the underlying true log price process $X_i(t)$ in continuous time, we observe $Y_i(t_{i\ell})$, the high-frequency noisy observations of $X_i(\cdot)$ at times $t_{i\ell}$, $\ell = 1, \ldots, n_i$, $i = 1, \ldots, p$. When estimating covolatilities of multiple assets based on high-frequency data, we encounter a well-known nonsynchronized problem, which refers to the fact that transactions for different assets often

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

occur at distinct times, and the high-frequency prices of different assets are recorded at mismatched time points. We allow the observations $Y_i(t_{i\ell})$ to be non-synchronized (that is, $t_{i\ell} \neq t_{j\ell}$ for any $i \neq j$). Because of microstructure noise in the high-frequency data, the observed log price $Y_i(t_{i\ell})$ is a noisy version of the corresponding true log price $X_i(t_{i\ell})$. In this paper, for the sake of simplicity we assume

$$Y_i(t_{i\ell}) = X_i(t_{i\ell}) + \varepsilon_i(t_{i\ell}), \qquad i = 1, \ldots, p, \qquad \ell = 1, \ldots, n_i, \tag{2}$$

where $\varepsilon_i(t_{i\ell})$, $i = 1, \ldots, p$, $\ell = 1, \ldots, n_i$, are independent noises with mean zero, for each fixed i, $\varepsilon_i(t_{i\ell})$, $\ell = 1, \ldots, n_i$, are independent and identically distributed (i.i.d.) random variables with variance η_i, and $\varepsilon_i(\cdot)$ and $X_i(\cdot)$ are independent. Here we adopt the i.i.d noise assumption for mathematical simplicity. We may relax the i.i.d. assumption to correlated noises such as the case considered in Ait-Sahalia, Mykland, and Zhang (2011) and still obtain the same convergence rates as in Section 4.

Our goal is to provide a new consistent estimator of the integrated volatility matrix $\mathbf{\Gamma} = \left(\Gamma_{ij}\right)_{1 \leq i,j \leq p}$ based on noisy and nonsynchronized high-frequency observations $Y_i(t_{i\ell})$, $\ell = 1, \ldots, n_i$, $i = 1, \ldots, p$, when p is very large. Recently Wang and Zou (2010) provided one consistent estimator when both p and the sample size can go to infinity. However, their estimator has a slow convergence rate. In this paper we shall construct a new estimator of $\mathbf{\Gamma}$ that has a fast convergence rate, depending on the sample size with the optimal $1/4$-exponent.

3. LARGE VOLATILITY MATRIX ESTIMATION

To construct a good estimator for large volatility matrix $\mathbf{\Gamma} = \left(\Gamma_{ij}\right)_{1 \leq i,j \leq p}$ using noisy high-frequency financial data, we need to take care of three issues: (1) the nonsynchronized problem; (2) microstructure noise; and (3) the number of assets p can be larger than sample size. When p is small and fixed, there are already many estimators proposed in the literature that take care of the first two issues. See, e.g., Ait-Sahalia et al. (2010), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), Christensen et al. (2010), Griffin and Oomen (2011), Hayashi and Yoshida (2005), Zhang (2011), and the references therein. However, none of these estimators perform well when the number of assets p is large. We shall adopt the sparsity and thresholding idea to handle the issue of fast-growing dimension p.

3.1. Multiscale Realized Volatility Matrix

Let m be an integer, and $\boldsymbol{\tau} = \{\tau_r, r = 1, \ldots, m\}$, where $\tau_r, r = 1, \ldots, m$, are a pre-determined sampling frequency. For asset i, define previous-tick times

$$\tau_{i,r} = \max\{t_{i\ell} \leq \tau_r, \ell = 1, \ldots, n_i\}, \qquad r = 1, \ldots, m.$$

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

With $\boldsymbol{\tau}$ we define realized covolatility between assets i and j by

$$\hat{\Gamma}_{ij}(\boldsymbol{\tau}) = \sum_{r=2}^{m} \left[Y_i(\tau_{i,r}) - Y_i(\tau_{i,r-1})\right] \left[Y_j(\tau_{j,r}) - Y_j(\tau_{j,r-1})\right], \tag{3}$$

and the previous-tick realized volatility matrix by

$$\hat{\boldsymbol{\Gamma}}(\boldsymbol{\tau}) = \left(\hat{\Gamma}_{ij}(\boldsymbol{\tau})\right). \tag{4}$$

We usually select the predetermined sampling frequency $\boldsymbol{\tau}$ to be regular grids. For a fixed m, we have $K = [n/m]$ classes of nonoverlap regular grids,

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

where $k = 1, \ldots, K$, and n is the average sample size

$$n = \frac{1}{p} \sum_{i=1}^{p} n_i.$$

Figure 1 illustrates the regular grids for $n_1 = 13$, $n_2 = 11$, $n_3 = 10$, $n = 11$, $m = 5$, and $K = 2$. The two panels in Figure 1 correspond to the cases of $k = 1$ and $k = 2$, respectively.

For each sampling frequency $\boldsymbol{\tau}^k$, we use (3) to define $\hat{\Gamma}_{ij}\left(\boldsymbol{\tau}^k\right)$ and realized covolatility matrix $\hat{\boldsymbol{\Gamma}}\left(\boldsymbol{\tau}^k\right)$. We average K realized covolatility matrices $\hat{\boldsymbol{\Gamma}}\left(\boldsymbol{\tau}^k\right)$ and define an average realized volatility matrix

$$\hat{\Gamma}_{ij}^{K} = \frac{1}{K} \sum_{k=1}^{K} \hat{\Gamma}_{ij}\left(\boldsymbol{\tau}^k\right), \qquad \hat{\boldsymbol{\Gamma}}^{K} = \left(\hat{\Gamma}_{ij}^{K}\right) = \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\Gamma}}\left(\boldsymbol{\tau}^k\right). \tag{6}$$

Wang and Zou (2010) defined an average realized volatility matrix (ARVM) estimator by adjusting the diagonal elements of $\hat{\boldsymbol{\Gamma}}^{K}$ for some bias corrections. The ARVM estimator is a two-scale estimator and has slow convergence rates with respect to sample size. To improve the ARVM estimator, we define our multiscale realized volatility matrix (MSRVM) estimator as

$$\hat{\boldsymbol{\Gamma}} = \sum_{m=1}^{M} a_m \hat{\boldsymbol{\Gamma}}^{K_m} + \zeta \left(\hat{\boldsymbol{\Gamma}}^{K_1} - \hat{\boldsymbol{\Gamma}}^{K_M}\right), \tag{7}$$

where

$$K_m = m + N, \qquad a_m = \frac{12(m+N)(m - M/2 - 1/2)}{M\left(M^2 - 1\right)}, \qquad \zeta = \frac{(M+N)(N+1)}{(n+1)(M-1)}. \tag{8}$$

We take M and N to be of the exact order $\sqrt{n}$. Note that when $p = 1$ our estimator becomes the multiscale realized volatility estimator proposed in Zhang (2006) and used by Fan and Wang (2007) for a single asset based on noisy high-frequency data.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

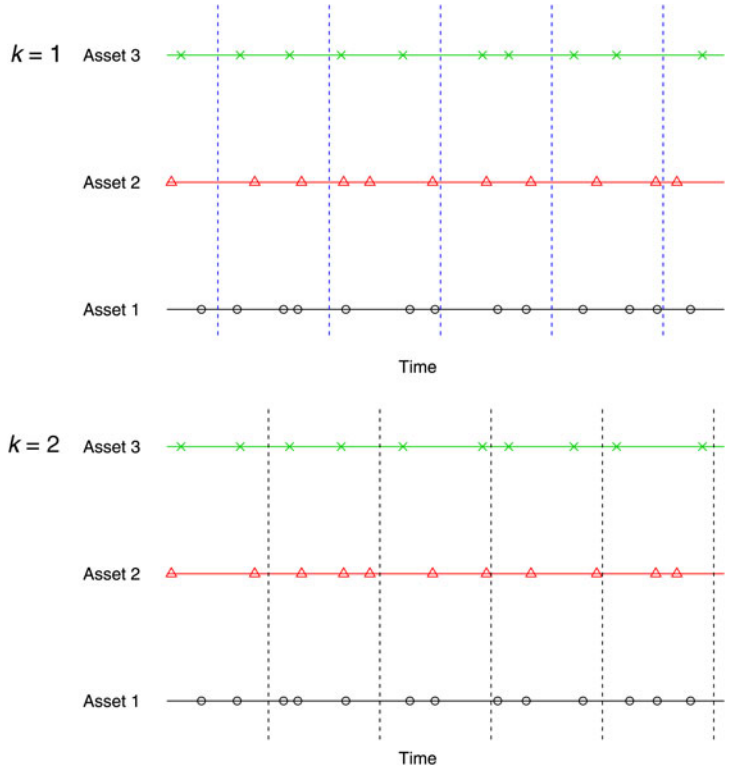


FIGURE 1. An illustration of regular grids with $K = 2$. Panels from up to down are the grids with $k = 1$ and $k = 2$, respectively.

3.2. Regularized Estimation of a Large Sparse Volatility Matrix

For a relatively small number of assets, p is much smaller than sample sizes, and all of the existing realized volatility matrix estimators (including our MSRVM estimator $\hat{\mathbf{\Gamma}}$) are consistent estimators of $\mathbf{\Gamma}$. However, they perform poorly when p is large. In fact, Wang and Zou (2010) and Tao et al. (2011) show that when both p and n go to infinity, all the existing realized volatility matrix estimators are inconsistent. The inconsistency indicates that for large p, the eigenvalues and eigenvectors of the volatility matrix estimators are far from those corresponding to $\mathbf{\Gamma}$. In order to consistently estimate $\mathbf{\Gamma}$ when p is very large, we have to impose some sparse structure on $\mathbf{\Gamma}$ and to regularize $\hat{\mathbf{\Gamma}}$.

There are different ways to impose sparse structures on large square matrices. As in Wang and Zou (2010) and Tao et al. (2011), we assume that $\mathbf{\Gamma}$ satisfies the sparse condition

$$\sum_{j=1}^{p} |\Gamma_{ij}|^{\delta} \le \Theta\, \pi(p), \qquad i = 1, \ldots, p, \qquad \mathbb{E}[\Theta] \le C, \tag{9}$$

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

where $0 \le \delta < 1$, $\pi(p)$ is a deterministic function of p that grows very slowly in p, Θ is a positive random variable, and C is a positive constant.

Remark 1. As the number of parameters to be estimated in volatility matrix $\mathbf{\Gamma}$ is of order p^2, for p is comparable to sample size, we can not estimate all these parameters consistently. The sparse modeling (9) makes consistent estimation of $\mathbf{\Gamma}$ possible for the large p scenario. Intuitively, a sparse matrix means that only a relatively small percent of elements in each row (or column) have significantly large magnitude (and thus are important). Now $\pi(p)$ controls the number of important elements that may grow with p, and typically we may take $\pi(p)$ to be 1 or $\log p$. Here δ calibrates the magnitude level of elements as significantly large. For example, $\delta = 0$ in (9) means that each row of $\mathbf{\Gamma}$ has at most $\Theta\pi(p)$ number of nonzero elements. Decay matrix $\mathbf{\Gamma}$ with $|\Gamma_{ij}| \le C|i-j|^{\alpha}$ corresponds to a special case of sparsity condition (9) with $\delta = 1/(\alpha+1)$ and $\pi(p) = \log p$ or $1/(\alpha+1) < \delta < 1$ and $\pi(p) = 1$.

Remark 2. Sparse modeling is widely used in scientific studies such as signal and image processing, remote sensing, and high-dimensional statistics. Sparse matrices include block diagonal matrices, matrices with decay elements from diagonal, matrices with a relatively small number of nonzero elements in each row or column, and matrices obtained by randomly permuting rows and columns of the above matrices. We can improve sparsity by transformations. For example, we may consider important economic factors and some known transformations like Fourier and wavelet transformations. After we sort out important factors and/or take some specific transformations, it is often the case that the volatility matrix resulting from the transformation of $\mathbf{\Gamma}$ is very sparse. In financial econometrics the generalized autoregressive conditionals heteroskedasticity (GARCH) modeling of large volatility matrices is usually to reduce large volatility matrices into a sequence of smaller matrices and transform the very high-dimensional model into a sequence of univariate or low-dimensional models (Palandri, 2009; Engle and Sheppard, 2001). The GARCH approaches correspond to a special case of sparsity where large volatility matrices are modeled as block diagonal matrices. General sparse modeling may be very useful for the study of high-dimensional economic and financial problems.

Given $\mathbf{\Gamma}$ satisfying the sparsity condition (9), its important elements are those whose absolute values are above a certain threshold. Thus we threshold the MSRVM estimator $\hat{\mathbf{\Gamma}}$ by retaining its elements with absolute values exceeding some given threshold value and replace others by zero; that is, we define

$$\tilde{\mathbf{\Gamma}} = \mathcal{T}_{\varpi}[\hat{\mathbf{\Gamma}}] = \left(\hat{\Gamma}_{ij} 1\left(|\hat{\Gamma}_{ij}| \ge \varpi\right)\right), \tag{10}$$

where ϖ is called threshold. The (i, j)th element of $\mathcal{T}_{\varpi}[\hat{\mathbf{\Gamma}}]$ is equal to $\hat{\Gamma}_{ij}$ if its absolute value is greater than or equal to ϖ and zero otherwise. The threshold value ϖ will affect the convergence rate, and the optimal ϖ will be given in Theorem 2.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

Similar to other existing covolatility matrix estimators, we can not guarantee the positive definite property for our estimator in finite samples. However, when both the sample sizes and the number of assets go to infinity, our estimator is asymptotically positive definite.

To summarize, our final estimator of $\mathbf{\Gamma}$ is the *threshold MSRVM* estimator given by equations (7), (8), and (10). As we show in the next section, our estimator is not only consistent but also converges to $\mathbf{\Gamma}$ at a rate faster than that of the estimators proposed in Wang and Zou (2010) and Tao et al. (2011), which are $\mathcal{T}_{\varpi}\left[\tilde{\mathbf{\Gamma}}^{K_m}\right]$ with $\tilde{\mathbf{\Gamma}}^{K_m}$ defined by (14) in Section 6.

4. ASYMPTOTIC THEORY

Let $\mathbf{x} = (x_1, \ldots, x_p)^T$ be a p-dimensional vector and $\mathbf{U} = (U_{ij})$ a p-by-p matrix. Define the ℓ_d-norms,

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

The matrix norm $\|\mathbf{U}\|_2$ is the square root of the largest eigenvalue of $\mathbf{U}\mathbf{U}^T$, and

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

For a symmetric $\mathbf{U}$, its matrix norm is equal to the largest absolute eigenvalue, and $\|\mathbf{U}\|_2 \le \|\mathbf{U}\|_1 = \|\mathbf{U}\|_\infty$. We consider ℓ_d-norm for $d = 1, 2, \infty$ in the paper.

We need the following technical conditions to establish the asymptotic theory.

Assumption A1. For some $\beta \ge 2$,

$$\max_{1\le i\le p} \max_{0\le t\le 1} \mathbb{E}\left[|\gamma_{ii}(t)|^\beta\right] < \infty, \qquad \max_{1\le i\le p} \max_{0\le t\le 1} \mathbb{E}\left[|\mu_i(t)|^{2\beta}\right] < \infty,$$

$$\max_{1\le i\le p} \mathbb{E}\left[|\varepsilon_i(t_{i\ell})|^{2\beta}\right] < \infty.$$

Assumption A2. There exist constants C_1 and C_2 such that

$$\max_{1\le i\le p} \frac{n_i}{n} \le C_1, \qquad \max_{1\le i\le p} \max_{1\le \ell\le n_i} |t_{i\ell} - t_{i,\ell-1}| \le C_2/n,$$

where $n = \left(n_1 + \cdots + n_p\right)/p$.

Remark 3. Assumption A1 imposes moment conditions on the price process and microstructure noise to derive the asymptotic theory. Also, we need some condition on sampling frequencies of the data in order to establish the asymptotic theory. Assumption A2 imposes conditions on data sampling frequencies over p assets. It essentially requires no "hole" in the entire observation time interval. If there is a "hole" in the data, no methods can consistently estimate the volatility

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

matrix over the hole. The second inequality in Assumption A2 implies that for all $1 \le i \le p$, $\frac{n_i}{n} \ge 1/C_2$. Since the constants C_1 and $1/C_2$ may differ substantially, the condition does not force all assets to sample at approximately equal rate. For example, if $C_1 C_2 = 20$, some assets may be sampled 20 times more often than other assets. Assumption A2 ensures that the data are observed at frequencies for which the gaps between adjacent time points are of order n^{-1}. Since K_m used in (7) are of order $n^{-1/2}$, from the definitions of (5)–(7) we see that for each of the M presampling grids, the gap between two consecutive grid points is equal to $1/K_m$, which is of exact order $n^{-1/2}$, for $m = 1, \ldots, M$. As $n^{-1/2}/n^{-1} = n^{1/2} \to \infty$, the selected presampling grids are an order of magnitude coarser than the observed data, and thus there are always some observations between any two consecutive grid points in the selected presampling grids. Assumption A2 can be relaxed. For example, we may allow n_i to vary within some powers of n and allow $\max_{i,\ell} |t_{i\ell} - t_{i,\ell-1}|$ to be a order of a power of $1/n$, and then we select K_m accordingly to construct a volatility matrix estimator and develop associated asymptotic theory. The important point is that we need to select presampling grids an order of magnitude coarser than the observed data.

We have the following results for estimating a large volatility matrix using noisy high-frequency data.

THEOREM 1. *Under Models (1)–(2) and Assumption A1–A2, the MSRVM estimator* $\hat{\mathbf{\Gamma}} = (\hat{\Gamma}_{ij})$ *given by (7)–(8) satisfies*

$$\mathbb{E}\left[\left|\hat{\Gamma}_{ij} - \Gamma_{ij}\right|^{\beta}\right] \le C n^{-\beta/4}, \qquad i, j = 1, \ldots, p, \tag{11}$$

where C is a generic constant free of n and p.

THEOREM 2. *Under Models (1)–(2), Assumption A1–A2, and sparsity (9), the threshold MSRVM estimator* $\mathcal{T}_{\varpi}(\hat{\mathbf{\Gamma}})$ *defined in (10) satisfies*

$$\mathbb{E}\left[\left\|\mathcal{T}_{\varpi}(\hat{\mathbf{\Gamma}}) - \mathbf{\Gamma}\right\|_2\right] \le \mathbb{E}\left[\left\|\mathcal{T}_{\varpi}(\hat{\mathbf{\Gamma}}) - \mathbf{\Gamma}\right\|_1\right] = O\left(\pi(p)\varpi^{1-\delta}\right),$$

where the threshold ϖ is of order $n^{-1/4} p^{2/\beta} h_{n,p}$, and $h_{n,p}$ is any sequence going to infinity arbitrarily slowly such as $h_{n,p} = \log\log(n \wedge p)$.

Remark 4. Theorem 1 indicates that the convergence rate for each element of the MSRVM estimator $\hat{\mathbf{\Gamma}}$ is $n^{-1/4}$, which is the optimal convergence rate for estimating each element of $\mathbf{\Gamma}$ (see Gloter and Jacod, 2001; Reiss, 2011). Note that, due to the contamination of high-frequency data by microstructure noise, this optimal convergence rate $n^{-1/4}$ is slower than the usual $n^{-1/2}$ rate.

Remark 5. As $h_{n,p}$ and $\pi(p)$ are slow growth factors, the convergence rate in Theorem 2 is nearly equal to $\left[n^{-1/4} p^{2/\beta}\right]^{1-\delta}$, which goes to zero when p grows more slowly than $n^{\beta/8}$. Assumption A1 imposes finite 2βth moments on the microstructure noise, drift, and diffusion covariance of log price $\mathbf{X}(t)$. Since it is always assumed that financial data have some finite moments, it is realistic to

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

assume Assumption A1 with reasonably large β. Thus, with p being allowed to be close to $n^{\beta/8}$ and β reasonably large, we can consistently estimate $\mathbf{\Gamma}$ by $\mathcal{T}_{\varpi}(\hat{\mathbf{\Gamma}})$ for p close to or even larger than n.

Remark 6. Wang and Zou (2010) and Tao et al. (2011) proposed their estimators as $\mathcal{T}_{\varpi}[\tilde{\mathbf{\Gamma}}^{K_m}]$ with $\tilde{\mathbf{\Gamma}}^{K_m}$ given by (14) in Section 5, and showed that they achieve the convergence rate of $\pi(p)\left[n^{-1/6}\,p^{2/\beta}\right]^{1-\delta}$, which is much slower than the convergence rate derived in Theorem 2 for our threshold MSRVM estimator.

Remark 7. For estimating the large sparse covariance matrix of Gaussian data, Bickel and Levina (2008) constructed a threshold estimator that can achieve convergence rate $\pi(p)\left[n^{-1/2}\log^{1/2}p\right]^{1-\delta}$, and Cai and Zhou (2012, 2013) showed that such a convergence rate is optimal. The convergence rate for the Gaussian data increases in matrix size p through $\log^{1/2}p$ and sample size via $n^{-1/2}$, while the convergence rate in Theorem 2 grows with n through $n^{-1/4}$ and p through a power of p. The slower convergence rate here in both p and n is due to the intrinsic complexity of our problem. The log p factor in the convergence rate of covariance matrix estimation is attributed to Gaussianity imposed on the observed data. In our setup, observations $Y_i(t_{i\ell})$ from model (2) have random sources from both microstructure noise $\varepsilon_i(t_{i\ell})$ and true log price $\mathbf{X}(t)$ given by model (1). First, as we have discussed in Remark 4, because of microstructure noise in the data, the optimal rate depends on n through $n^{-1/4}$ instead of $n^{-1/2}$ for covariance matrix estimation; second, log price $\mathbf{X}(t)$ has finite moments but does not obey Gaussianity or sub-Gaussianity for common price and volatility models. Because we impose only realistic finite moment conditions in Assumption A1, the obtained convergence rate in Theorem 2 depends on p through a power of p instead of log p for the Gaussian case. These facts lead us to conjecture that the convergence rate in Theorem 2 is the optimal convergence rate with respect to both n and p for the large volatility matrix estimation problem in our setup.

5. NUMERICAL STUDIES

We conducted simulations to check the performances of the proposed estimators and compare them with the ARVM-based estimators for finite samples. We simulated the true log price $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^T$ at $t_\ell = \ell/n$, $\ell = 1, \ldots, n$, from model (1) with $\boldsymbol{\mu}_t = 0$ and volatility matrix $\boldsymbol{\sigma}_t$ as a Cholesky decomposition of

$$\boldsymbol{\gamma}(t) = \big(\gamma_{ij}(t)\big), \qquad \gamma_{ij}(t) = \tau_i \tau_j \kappa^{|i-j|}, \tag{12}$$

where $\{\tau_i,\ i = 1, \ldots, p\}$ are independently simulated from a uniform distribution on $(0, 1)$, and κ is taken to be 0.5. We generated synchronized noisy observations $Y_i(t_\ell)$ from model (2) by adding mean zero normal random errors $\varepsilon_i(t_\ell)$ to the simulated log price $X_i(t_\ell)$, $\ell = 1, \ldots, n$, where for the ith asset, the random errors $\varepsilon_i(t_{i\ell})$ have standard deviation $\theta\tau_i$, τ_i are given by (12), and θ is the relative

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

noise level with range selected from 0 to 0.6 in the simulation study. We used the simulated data $Y_i(t_\ell)$ to compute the MSRVM estimator and the threshold MSRVM estimator given in Section 3, as well as the ARVM estimator and the threshold ARVM estimator defined in Wang and Zou (2010).

In the simulation study we fixed $n = 200$ and chose two values of $p = 3,100$. We repeated the whole simulation procedure 200 times. The mean ℓ_2 error (ME) of each matrix estimator is computed by averaging ℓ_2-norms of the differences between the estimator and $\mathbf{\Gamma}$ over 200 repetitions. We used the MEs to evaluate the performances of the estimators. In the simulation study we selected values for K in the ARVM estimator, M and N in the MSRVM estimator, and threshold ϖ in the threshold ARVM estimator and the threshold MSRVM estimator by minimizing their respective MEs.

We started with simulation results for small $p = 3$. Figure 2 plots the MEs against noise level for the MSRVM estimator, the threshold MSRVM estimator, the ARVM estimator, and the threshold ARVM estimator. It shows that at the low noise level, the ARVM estimator performs better than the MSRVM estimator. As the noise level increases, the ME of the ARVM estimator increases dramatically and quickly exceeds the ME of the MSRVM estimator, which decreases initially and then increases slightly. The simulations also show that at all noise levels, the MEs of the ARVM and MSRVM estimators are very close to those of the corresponding threshold estimators. The findings can be heuristically explained as follows. The higher ME of the MSRVM estimator at very low noise level is due to the fact that for the noiseless case, the best estimator is the simple realized volatility, which is a one-scale estimator, and the purpose of the two-scale and multiscale schemes is for noise reduction. At the very low noise level, where there is not much noise to reduce, the complicated multiscale scheme in the MSRVM estimator produces larger bias than the two-scale design in the ARVM estimator,

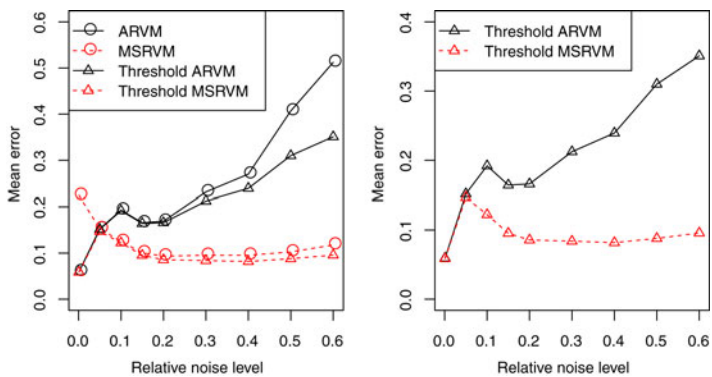


FIGURE 2. The ME plot for the four estimators with $p = 3$. The left panel is the ME curves for all four estimators, and the right panel is the ME curves for the threshold MSRVM and threshold ARVM estimators.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

and thus the MSRVM estimator has larger ME than the ARVM estimator. On the other hand, at the higher noise level, the multiscale MSRVM estimator is more effective in reducing noise and thus yields significantly smaller ME than the ARVM estimator. From the plot we see that thresholding does not improve the estimators. The reason is that $p = 3$ is very small relative to sample size $n = 200$, for which thresholding is not needed.

For the scenario of large p, we reported the simulation results for $p = 100$. Figure 3 plots the MEs of the four estimators against noise level, and Tables 1 and 2 provide their MEs and average values for (K, M, N, ϖ) along with the

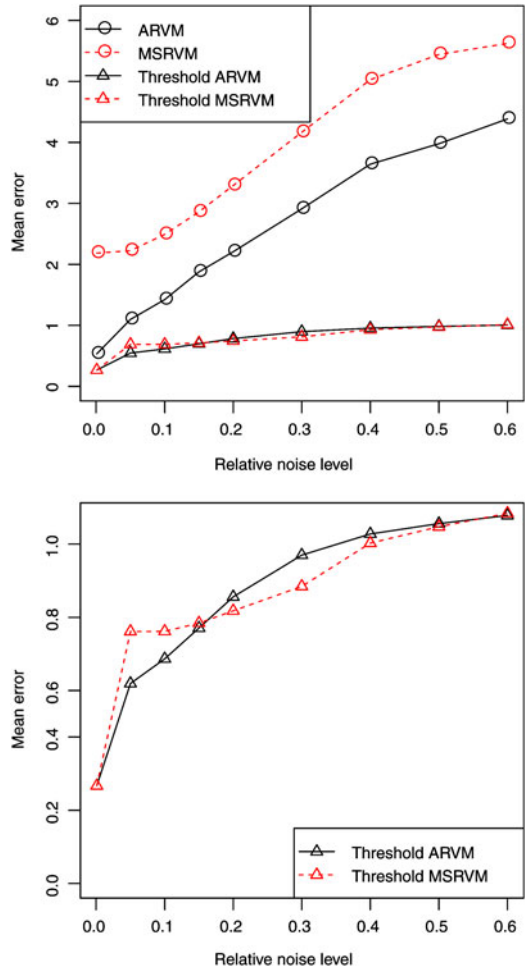


FIGURE 3. The ME plot for the four estimators with $p = 100$. The upper panel is the ME curves for the four estimators, and the lower panel is the ME curves for threshold MSRVM and threshold ARVM estimators.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

TABLE 1. Simulation results for the MSRVM estimator ($\hat{\mathbf{\Gamma}}$) and the threshold MSRVM estimator ($\tilde{\mathbf{\Gamma}}$) with $p = 100$

θ	ME of $\hat{\mathbf{\Gamma}}$	ME of $\tilde{\mathbf{\Gamma}}$	$\bar{\varpi}$	$\bar{M}$	$\bar{N}$
0.001	2.186 (0.031)	0.267 (0.003)	0.426 (0.009)	5.0 (0.0)	2.0 (0.0)
0.05	2.225 (0.038)	0.689 (0.009)	0.440 (0.008)	5.3 (0.1)	2.0 (0.0)
0.1	2.493 (0.064)	0.689 (0.011)	0.546 (0.011)	5.9 (0.2)	2.0 (0.0)
0.15	2.861 (0.077)	0.711 (0.013)	0.665 (0.011)	7.5 (0.3)	2.3 (0.1)
0.2	3.293 (0.059)	0.745 (0.012)	0.720 (0.012)	10.0 (0.2)	2.5 (0.1)
0.3	4.167 (0.079)	0.812 (0.019)	0.772 (0.012)	15.4 (0.4)	3.9 (0.2)
0.4	5.029 (0.095)	0.930 (0.033)	0.845 (0.010)	20.2 (0.4)	6.2 (0.4)
0.5	5.443 (0.109)	0.975 (0.038)	0.874 (0.009)	23.5 (0.4)	8.1 (0.4)
0.6	5.624 (0.122)	1.011 (0.041)	0.885 (0.008)	25.4 (0.4)	9.7 (0.5)

Notes: The θ column is the relative noise level for the microstructure noise, the $\bar{M}$ and $\bar{N}$ columns are the respective average values of M and N used in $\hat{\mathbf{\Gamma}}$ over 200 repetitions, and the $\bar{\varpi}$ column is the average value of threshold ϖ used in $\tilde{\mathbf{\Gamma}}$ over 200 repetitions. The values in parentheses represent the corresponding standard errors.

corresponding standard errors. The basic findings are that thresholding significantly improves both the MSRVM and ARVM estimators. Figure 3 shows that except for the very small noise level case, the MEs of the MSRVM and ARVM estimators are much larger than the threshold counterparts. Both the threshold ARVM estimator and the threshold MSRVM estimator perform very well, compared with the MSRVM and ARVM estimators. The threshold ARVM estimator performs a little bit better than the threshold MSRVM estimator at very low noise levels, and the threshold MSRVM estimator has slightly smaller ME than the

TABLE 2. Simulation results for the ARVM estimator and the threshold ARVM estimator with $p = 100$

θ	ME of ARVM	ME of threshold ARVM	$\bar{\varpi}$	$\bar{K}$
0.001	0.533 (0.007)	0.267 (0.003)	0.102 (0.001)	1.0 (0.0)
0.05	1.098 (0.021)	0.546 (0.004)	0.188 (0.004)	5.1 (0.1)
0.1	1.423 (0.028)	0.614 (0.006)	0.260 (0.004)	6.5 (0.1)
0.15	1.878 (0.034)	0.698 (0.006)	0.342 (0.005)	9.9 (0.2)
0.2	2.208 (0.036)	0.783 (0.007)	0.412 (0.007)	13.1 (0.3)
0.3	2.911 (0.053)	0.897 (0.005)	0.588 (0.010)	19.8 (0.5)
0.4	3.643 (0.078)	0.955 (0.005)	0.772 (0.010)	27.1 (0.8)
0.5	3.978 (0.078)	0.983 (0.007)	0.828 (0.009)	34.2 (0.8)
0.6	4.385 (0.093)	1.006 (0.010)	0.865 (0.008)	42.9 (0.7)

Notes: The θ column is the relative noise level for the microstructure noise, the $\bar{K}$ column is the average value of K in (6) used for the ARVM estimator over 200 repetitions, and the $\bar{\varpi}$ column is the average value of threshold ϖ used in the threshold ARVM estimator over 200 repetitions. The values in parentheses represent the corresponding standard errors.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

threshold ARVM estimator at higher noise levels. The behaviors of the threshold ARVM estimator and the MSRVM estimator with large $p = 100$ are quite similar to those of the ARVM estimator and the MSRVM estimator with small $p = 3$. While confirming the findings on ME, Tables 1 and 2 reveal that as noise level increases, MEs and the average values of (K, M, N, ϖ) all increase. The higher the noise level is, the harder the estimation problem becomes. As a result, the estimators have larger MEs, and naturally we need to use larger (K, M, N) in the ARVM estimator and the MSRVM estimator to reduce noise and select bigger threshold ϖ to better balance bias and variance.

We further studied the performances of the proposed MSRVM estimator and the threshold MSRVM estimator in terms of the whole range of eigenvalues. Figure 4 displays the 100 eigenvalues of volatility matrix $\mathbf{\Gamma}$, the average eigenvalue curve for each of the MSRVM estimator, the threshold MSRVM estimator, and the threshold ARVM estimator for $\theta = 0.2$ and $p = 100$, where each average eigenvalue curve represents 100 average eigenvalues over 200 repetitions for the corresponding estimator. This figure shows that while the average eigenvalues of the MSRVM estimator are far off from the true eigenvalues of $\mathbf{\Gamma}$ at the two extremes, the average eigenvalues of the threshold MSRVM estimator are very close to the true eigenvalues. Furthermore, between the two threshold estimators, the threshold MSRVM estimator has eigenvalues much closer to the true eigenvalues than the threshold ARVM estimator. The conclusions reinforce the ME-based performance findings for the MSRVM estimator, the threshold MSRVM estimator, and the threshold ARVM estimator.

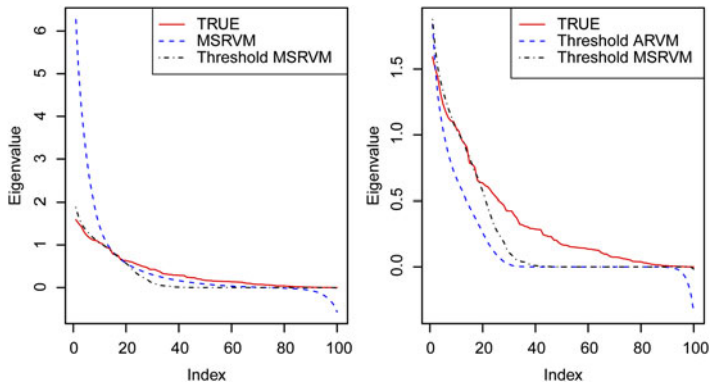


FIGURE 4. The eigenvalue plot of $\mathbf{\Gamma}$, the MSRVM estimator, the threshold MSRVM estimator, and the threshold ARVM estimator with noise level $\theta = 0.2$ and $p = 100$. The left panel is the plot of the 100 average eigenvalues for each of MSRVM and threshold MSRVM estimators over 200 repetitions. The right panel is the plot of the 100 average eigenvalues for each of threshold ARVM and threshold MSRVM estimators over 200 repetitions.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

6. PROOFS

Denote by C a generic constant whose value is free of n and p and may change from appearance to appearance.

Proof of Theorem 1. Define

$$\hat{\boldsymbol{\eta}} = \left(\hat{\eta}_{ij}\right) = \operatorname{diag}\left(\hat{\eta}_1, \ldots, \hat{\eta}_p\right), \qquad \hat{\eta}_i = \frac{1}{2n_i} \sum_{\ell=2}^{n_i} \left[Y_i\left(t_{i,\ell}\right) - Y_i\left(t_{i,\ell-1}\right)\right]^2, \tag{13}$$

$$\tilde{\boldsymbol{\Gamma}}^{K_m} = \hat{\boldsymbol{\Gamma}}^{K_m} - 2\frac{n - K_m + 1}{K_m}\hat{\boldsymbol{\eta}}. \tag{14}$$

Then $\hat{\eta}_i$ is a consistent estimator of the variance, $\eta_i = \operatorname{Var}(\varepsilon_i(t_{i\ell}))$, of the microstructure noise for the ith asset, and $\tilde{\boldsymbol{\Gamma}}^{K_m}$ are the ARVM estimators given in Wang and Zou (2010). Applying Theorem 1 of Wang and Zou to $\tilde{\boldsymbol{\Gamma}}^{K_m}$, we have

$$\mathbb{E}\left(|\tilde{\Gamma}_{ij}^{K_m} - \Gamma_{ij}|^\beta\right) \le C\left[\left(K_m n^{-1/2}\right)^{-\beta} + K_m^{-\beta/2} + \left(n/K_m\right)^{-\beta/2} + K_m^{-\beta} + n^{-\beta/2}\right]. \tag{15}$$

From the definition of $\hat{\boldsymbol{\Gamma}}$ given in (7)–(8) we obtain

$$\begin{aligned}\hat{\boldsymbol{\Gamma}} = &\sum_{m=1}^{M} a_m \tilde{\boldsymbol{\Gamma}}^{K_m} + \zeta\left(\tilde{\boldsymbol{\Gamma}}^{K_1} - \tilde{\boldsymbol{\Gamma}}^{K_M}\right) \\ &+ 2\left[\sum_{m=1}^{M} a_m \frac{n - K_m + 1}{K_m} + \zeta\left(\frac{n - K_1 + 1}{K_1} - \frac{n - K_M + 1}{K_M}\right)\right]\hat{\boldsymbol{\eta}},\end{aligned}$$

and therefore,

$$\begin{aligned}\mathbb{E}|\hat{\Gamma}_{ij} - \Gamma_{ij}|^\beta \le C\Bigg\{&\mathbb{E}\left|\sum_{m=1}^{M} a_m\left(\tilde{\Gamma}_{ij}^{K_m} - \Gamma_{ij}\right)\right|^\beta + \mathbb{E}\left|\zeta\left(\tilde{\Gamma}_{ij}^{K_1} - \tilde{\Gamma}_{ij}^{K_M}\right)\right|^\beta \\ &+ 2\mathbb{E}\left|\left[\sum_{m=1}^{M} a_m \frac{n - K_m + 1}{K_m} + \zeta\left(\frac{n - K_1 + 1}{K_1} - \frac{n - K_M + 1}{K_M}\right)\right]\hat{\eta}_{ij}\right|^\beta\Bigg\} \\ = &\ C\,(\mathrm{I} + \mathrm{II} + \mathrm{III}).\end{aligned} \tag{16}$$

We prove the theorem by showing that I, II, and III are all of order $n^{-\beta/2}$ below.

For Part III on the right-hand side of (16), from the definitions of a_m, K_m, and ζ in (8), we evaluate the coefficient of $\hat{\eta}_{ij}$ in (16),

$$\begin{aligned}&\sum_{m=1}^{M} a_m \frac{n - K_m + 1}{K_m} + \zeta\left(\frac{n - K_1 + 1}{K_1} - \frac{n - K_M + 1}{K_M}\right) \\ &\quad = -1 + \frac{(M+N)(N+1)}{(n+1)(M-1)}(n+1)\left(\frac{M-1}{(N+1)(N+M)}\right) = 0,\end{aligned}$$

and thus $III = 0$.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

We consider Part I on the right-hand side of (16). Define

$$\mathbf{U}^{K_m} = \mathbf{G}^{K_m}(1) - 2\frac{n - K_m + 1}{K_m}\hat{\boldsymbol{\eta}}, \tag{17}$$

$$\mathbf{R}^{K_m} = \left[\mathbf{G}^{K_m}(2) + \mathbf{G}^{K_m}(3)\right] + \left[\mathbf{V}^{K_m} - \boldsymbol{\Gamma}\right] + \left[\mathbf{H}^{K_m}(1) + \cdots + \mathbf{H}^{K_m}(8)\right], \tag{18}$$

where $\mathbf{G}^{K_m}(1), \mathbf{G}^{K_m}(2), \mathbf{G}^{K_m}(3)$, $\mathbf{V}^{K_m}$, and $\mathbf{H}^{K_m}(1), \ldots, \mathbf{H}^{K_m}(8)$ are the same as those in the proof of Theorem 1 of Wang and Zou (2010, Sec. 7.1). Then Part I can be written as

$$\mathrm{I} = \mathbb{E}\left|\sum_{m=1}^{M} a_m \mathbf{U}^{K_m} + \sum_{m=1}^{M} a_m \mathbf{R}^{K_m}\right|^{\beta} \le C\left(\mathbb{E}\left|\sum_{m=1}^{M} a_m \mathbf{U}^{K_m}\right|^{\beta} + \mathbb{E}\left|\sum_{m=1}^{M} a_m \mathbf{R}^{K_m}\right|^{\beta}\right). \tag{19}$$

From (17) we can bound the first term on the right-hand side of (19) as

$$\begin{aligned}
\mathbb{E}\left|\sum_{m=1}^{M} a_m U_{ij}^{K_m}\right|^{\beta} &\le \frac{C}{M^{1-\beta/2}} \sum_{m=1}^{M} \mathbb{E}\left|a_m U_{ij}^{K_m}\right|^{\beta} \\
&\le \frac{C}{M^{1-\beta/2}} \sum_{m=1}^{M} \frac{12^{\beta}(m+N)^{\beta}\,|m - M/2 - 1/2|^{\beta}}{[M(M^2-1)]^{\beta}} \frac{n^{\beta/2}}{(m+N)^{\beta}} \\
&\le \frac{C}{M^{1-\beta/2}} n^{\beta/2} M^{\beta+1-3\beta} \le C n^{-\beta/4},
\end{aligned} \tag{20}$$

where the last inequality is due to $M \sim n^{1/2}$. To derive the second inequality in the above array, we use the definitions of a_m and K_m in (8), $M = N \sim n^{1/2}$. Furthermore, because $\mathbf{U}^{K_m}$ are martingales, we apply the Burkholder inequality (Chow and Teicher, 1997, Sec. 11.2) to $\mathbf{U}^{K_m}$ and obtain

$$\mathbb{E}|U_{ij}^{(K_m)}|^{\beta} \le C\left(n/K_m^2\right)^{\beta/2}.$$

See also Wang and Zou (2010, Prop. 1).

From the definitions of a_m and K_m in (8), we can easily verify that $\sum_{m=1}^{M} a_m = 1$, $\sum_{m=1}^{M} a_m/K_m = 0$, and $\sum_{m=1}^{M} |a_m| \sim \frac{3}{2} + \frac{3N}{M} = 9/2$. By (18) we establish an upper bound for the second term on the right-hand side of (19) below,

$$\begin{aligned}
\mathbb{E}\left|\sum_{m=1}^{M} a_m \mathbf{R}^{K_m}\right|^{\beta} \le \frac{C}{M^{1-\beta/2}} \Bigg(&\sum_{m=1}^{M} \mathbb{E}\left|a_m\left[G_{ij}^{K_m}(2) + G_{ij}^{K_m}(3)\right]\right|^{\beta} \\
&+ \sum_{m=1}^{M} \mathbb{E}\left|a_m\left[V_{ij}^{K_m} - \Gamma_{ij}\right]\right|^{\beta} \\
&+ \sum_{m=1}^{M} \mathbb{E}\left|a_m\left[H^{K_m}(1) + \ldots + H^{K_m}(8)\right]\right|^{\beta}\Bigg) \\
\le C\left(M^{-\beta/2} + (n/M)^{-\beta/2} + M^{-\beta} + n^{-\beta/2}\right) &\le C n^{-\beta/4},
\end{aligned} \tag{21}$$

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

where the second inequality is due to Propositions 1–3 in Wang and Zou (2010, Sec. 7), and the last inequality is a consequence of $M \sim n^{1/2}$.

Plugging (20) and (21) into (19), we conclude that $I \le C n^{-\beta/4}$.

Finally, we bound Part II on the right-hand side of (16) as

$$\begin{aligned}\mathbb{E}\left|\zeta\left(\tilde{\Gamma}_{ij}^{K_1}-\tilde{\Gamma}_{ij}^{K_M}\right)\right|^{\beta} &= |\zeta|^{\beta}\mathbb{E}\left|\left(\tilde{\Gamma}_{ij}^{K_1}-\Gamma_{ij}\right)-\left(\tilde{\Gamma}_{ij}^{K_M}-\Gamma_{ij}\right)\right|^{\beta}\\ &\le C\left(\frac{n}{M}\right)^{-\beta}\left(\mathbb{E}\left|\tilde{\Gamma}_{ij}^{K_1}-\Gamma_{ij}\right|^{\beta}+\mathbb{E}\left|\tilde{\Gamma}_{ij}^{K_M}-\Gamma_{ij}\right|^{\beta}\right)\\ &\le C\left(\frac{n}{M}\right)^{-\beta}\\ &\le C n^{-\beta/2},\end{aligned}$$

where in the above equation array of four lines, the inequality in the second line is from the definition of ζ in (8), and the inequality in the third line is due to the fact that K_1 and K_M are of order $n^{1/2}$ according to the definitions in (8), and Theorem 1 in Wang and Zou (2010) shows that $\mathbb{E}\left|\tilde{\Gamma}_{ij}^{K_1}-\Gamma_{ij}\right|^{\beta}$ and $\mathbb{E}\left|\tilde{\Gamma}_{ij}^{K_M}-\Gamma_{ij}\right|^{\beta}$ are bounded. ■

Proof of Theorem 2. With $\mathbf{\Gamma}=(\Gamma_{ij})$ and $\hat{\mathbf{\Gamma}}=\left(\hat{\Gamma}_{ij}\right)$, let $\tilde{\mathbf{\Gamma}}=\left(\tilde{\Gamma}_{ij}\right)=\mathcal{T}_{\varpi}\left(\hat{\mathbf{\Gamma}}\right)$, and $\tilde{\Gamma}_{ij}=\hat{\Gamma}_{ij}1\left(|\hat{\Gamma}_{ij}|\le\varpi\right)$. Define $d_{ij}=\left(\tilde{\Gamma}_{ij}-\Gamma_{ij}\right)1(A_{ij}^{c})$, $i,j=1,\ldots,p$, $\mathbf{D}=(d_{ij})$, where $A_{ij}=\left\{|\tilde{\Gamma}_{ij}-\Gamma_{ij}|\le 4\min\{|\Gamma_{ij},\varpi|\}\right\}$. With the definition of A_{ij} and $\mathbf{D}$, $\mathbb{E}\|\tilde{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_1$ can be bounded as

$$\mathbb{E}\|\tilde{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_1 \le \mathbb{E}\|\tilde{\mathbf{\Gamma}}-\mathbf{\Gamma}-\mathbf{D}\|_1+\mathbb{E}\|\mathbf{D}\|_1, \tag{22}$$

where we will show below that the first term on the right-hand side of (22) has the order of $\pi(p)\varpi^{1-\delta}$, and the second term is negligible.

Applying the Chebyshev inequality we have, for any fixed a,

$$P\left(|\hat{\Gamma}_{ij}-\Gamma_{ij}|\ge a\varpi\right)\le\frac{\mathbb{E}|\hat{\Gamma}_{ij}-\Gamma_{ij}|^{\beta}}{(a\varpi)^{\beta}}\le\frac{Cn^{-\beta/4}}{a^{\beta}n^{-\beta/4}p^{2}h_{n,p}^{\beta}}=Ca^{-\beta}p^{-2}h_{n,p}^{-\beta}, \tag{23}$$

where the second inequality is due to Theorem 1.

For the first term on the right-hand side of (22), conditional on the whole volatility process we obtain an upper bound on the conditional expectation,

$$\begin{aligned}\mathbb{E}\left(\|\tilde{\mathbf{\Gamma}}-\mathbf{\Gamma}-\mathbf{D}\|_1\middle|\mathbf{\Gamma}\right)&=\mathbb{E}\left(\sup_j\sum_i|\tilde{\Gamma}_{ij}-\Gamma_{ij}|1(A_{ij})\middle|\mathbf{\Gamma}\right)\\ &\le 4\mathbb{E}\left(\sup_j\sum_i\min\{|\Gamma_{ij},\varpi|\}\middle|\mathbf{\Gamma}\right)=4\sup_j\sum_i\min\{|\Gamma_{ij},\varpi|\}\end{aligned}$$

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

$$= 4\sup_j \sum_i |\Gamma_{ij}| 1(|\Gamma_{ij}| < \varpi) + 4\sup_j \sum_i \varpi 1(|\Gamma_{ij}| \geq \varpi)$$
$$\leq \Theta\pi(p)\varpi^{1-\delta} + \varpi \cdot \Theta\pi(p)\varpi^{-\delta} = 2\Theta\pi(p)\varpi^{1-\delta},$$

where the last inequality is due to Lemma 1 in Wang and Zou (2010, Sec. 6). Hence,

$$\mathbb{E}\|\tilde{\mathbf{\Gamma}} - \mathbf{\Gamma} - \mathbf{D}\|_1 = \mathbb{E}\left\{\mathbb{E}\left(\left\|\tilde{\mathbf{\Gamma}} - \mathbf{\Gamma} - \mathbf{D}\right\|_1 \middle| \mathbf{\Gamma}\right)\right\}$$
$$\leq \mathbb{E}\left(2\Theta\pi(p)\varpi^{1-\delta}\right) \leq C\pi(p)\varpi^{1-\delta}.$$

We consider the second term on the right-hand side of (22). Note that

$$\mathbb{E}\|\mathbf{D}\|_1 = \mathbb{E}\left(\sup_j \sum_i |d_{ij}|\right)$$
$$\leq \mathbb{E}\left(\sup_j \sum_i |d_{ij}| 1\left(\tilde{\Gamma}_{ij} = 0\right)\right) + \mathbb{E}\left(\sup_j \sum_i |d_{ij}| 1\left(\tilde{\Gamma}_{ij} = \hat{\Gamma}_{ij}\right)\right)$$
$$= I_1 + I_2. \tag{24}$$

Below we will show that both I_1 and I_2, are of order $\pi(p)\varpi^{1-\delta}$, and thus $\mathbb{E}\|\mathbf{D}\|_1$ has the desired order.

First we derive the order for I_1,

$$I_1 = \mathbb{E}\left(\sup_j \sum_i |\Gamma_{ij}| 1\left(|\Gamma_{ij}| > 4\varpi\right) 1\left(|\hat{\Gamma}_{ij}| < \varpi\right)\right)$$
$$\leq \mathbb{E}\left(\sup_j \sum_i |\hat{\Gamma}_{ij} - \Gamma_{ij}| 1\left(|\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq 3\varpi\right)\right) + \mathbb{E}\left(\sup_j \sum_i \varpi 1\left(|\Gamma_{ij}| \geq 4\varpi\right)\right)$$
$$\leq \sum_{i,j}\left(\mathbb{E}|\hat{\Gamma}_{ij} - \Gamma_{ij}|^\beta\right)^{1/\beta}\left(P\left(|\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq 3\varpi\right)\right)^{1-1/\beta}$$
$$+ \varpi\mathbb{E}\left\{\mathbb{E}\left(\sup_j \sum_i 1(|\Gamma_{ij}| \geq 4\varpi) \middle| \Gamma_{ij}\right)\right\}$$
$$\leq p^2 C n^{-1/4}\left(Cp^{-2}h_{n,p}^{-\beta}\right)^{1-1/\beta} + \varpi\mathbb{E}\left(4^{-\delta}\Theta\pi(p)\varpi^{-\delta}\right)$$
$$= Ch_{n,p}^{-\beta}\varpi + C\pi(p)\varpi^{1-\delta} = O\left(\pi(p)\varpi^{1-\delta}\right),$$

where in the above equation array of five lines, the inequality in the third line is due to the Hölder inequality, and the inequality in the fourth line is due to Theorem 1, inequality (23), and Lemma 1 in Wang and Zou (2010, Sec. 6).

Next we show that I_2 in (24) has the same order as

$$I_2 \leq \sum_{i,j}\mathbb{E}\left[|d_{ij}| 1\left(\tilde{\Gamma}_{ij} = \hat{\Gamma}_{ij}\right)\right]$$
$$= \sum_{i,j}\mathbb{E}\left[|\hat{\Gamma}_{ij} - \Gamma_{ij}| 1\left(A_{ij}^c\right)\right] \leq \sum_{i,j}\left(\mathbb{E}|\hat{\Gamma}_{ij} - \Gamma_{ij}|^\beta\right)^{1/\beta}\left(P\left(A_{ij}^c\right)\right)^{1-1/\beta}$$

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

$$\leq p^2 C n^{-1/4} \left(2Cp^{-2}h_{n,p}^{-\beta}\right)^{1-1/\beta}$$
$$= Ch_{n,p}^{-\beta}\varpi = o\left(\pi(p)\varpi^{1-\delta}\right),$$

where in the above equation array of four lines, the inequality in the second line is due to the Hölder inequality, and the inequality in the third line is due to Theorem 1 and the inequality

$$P\left(A_{ij}^c\right) \leq 2Cp^{-2}h_{n,p}^{-\beta}. \tag{25}$$

The rest of the proof is to show (25).

Let $A_1 = \left\{|\hat{\Gamma}_{ij}| \geq \varpi\right\}$. From the definition of $\tilde{\Gamma}_{ij}$ we have

$$|\tilde{\Gamma}_{ij} - \Gamma_{ij}| = |\Gamma_{ij}|1\left(A_1^c\right) + |\hat{\Gamma}_{ij} - \Gamma_{ij}|1\left(A_1\right).$$

Note that

$$A_1 = \left\{|\hat{\Gamma}_{ij}| \geq \varpi\right\} = \left\{|\hat{\Gamma}_{ij} - \Gamma_{ij} + \Gamma_{ij}| \geq \varpi\right\} \subset \left\{|\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq \varpi - |\Gamma_{ij}|\right\},$$
$$A_1^c = \left\{|\hat{\Gamma}_{ij}| < \varpi\right\} = \left\{|\hat{\Gamma}_{ij} - \Gamma_{ij} + \Gamma_{ij}| < \varpi\right\} \subset \left\{|\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq |\Gamma_{ij}| - \varpi\right\}.$$

An application of (23) leads to

$$P\left(A_1, |\Gamma_{ij}| < \varpi/4\right) \leq P\left(|\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq 3\varpi/4\right) \leq Cp^{-2}h_{n,p}^{-\beta},$$
$$P\left(A_1^c, |\Gamma_{ij}| > 2\varpi\right) \leq P\left(|\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq \varpi\right) \leq Cp^{-2}h_{n,p}^{-\beta},$$

and hence with probability at least $1 - Cp^{-2}h_{n,p}^{-\beta}$,

$$|\tilde{\Gamma}_{ij} - \Gamma_{ij}| = \begin{cases} |\Gamma_{ij}|, & |\Gamma_{ij}| < \varpi/4 \\ |\Gamma_{ij}| \text{ or } |\hat{\Gamma}_{ij} - \Gamma_{ij}|, & \varpi/4 \leq |\Gamma_{ij}| \leq 2\varpi \\ |\hat{\Gamma}_{ij} - \Gamma_{ij}|, & |\Gamma_{ij}| > 2\varpi. \end{cases} \tag{26}$$

Note that

$$4\min\{|\Gamma_{ij}|, \varpi\} = \begin{cases} 4|\Gamma_{ij}|, & |\Gamma_{ij}| < \varpi/4 \\ \min\{4\Gamma_{ij}, 4\varpi\} \geq \max\{|\Gamma_{ij}|, \varpi\}, & \varpi/4 \leq |\Gamma_{ij}| \leq 2\varpi \\ 4\varpi, & |\Gamma_{ij}| > 2\varpi. \end{cases} \tag{27}$$

Again, (23) implies that with probability at least $1 - Cp^{-2}h_{n,p}^{-\beta}$, we have $|\hat{\Gamma}_{ij} - \Gamma_{ij}| \leq \varpi$. Combining this result with (26) and (27), we conclude that with probability at least $1 - Cp^{-2}h_{n,p}^{-\beta}$, $|\tilde{\Gamma}_{ij} - \Gamma_{ij}| \leq 4\min\{\Gamma_{ij}, \varpi\}$, which proves (25). ■

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press

REFERENCES

Ait-Sahalia, Y., J. Fan, & D. Xiu (2010) High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* 105, 1504–1517.

Ait-Sahalia, Y., P. Mykland, & L. Zhang (2011) Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* 160, 190–203.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2011) Multivariate realized kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149–169.

Bickel, P.J. & E. Levina (2008) Covariance regularization by thresholding. *Annals of Statistics* 36, 2577–2604.

Cai, T. & H.H. Zhou (2012) Minimax estimation of large covariance matrices under ℓ_1-norm (with discussion). *Statistica Sinica* 22, 1319–1378.

Cai, T. & H.H. Zhou (2013) Optimal rates of convergence for sparse covariance matrix estimation. *Annals of Statistics*, in press.

Chow, Y.S. & H. Teicher (1997) *Probability Theory: Independence, Interchangeability, Martingales*, 3rd ed. Springer.

Christensen, K., S. Kinnebrock, & M. Podolskij (2010) Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics* 159, 116–133.

Engle, R.F. & K. Sheppard (2001) Theoretical and Empirical Properties of Dynamic Conditional Correlation Matrices. Working paper, Oxford University.

Fan, J. & Y. Wang (2007) Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of American Statistical Association* 102, 1349–1362.

Gloter, A. & J. Jacod (2001) Diffusions with measurement errors. I.: Local asymptotic normality. *ESAIM* 5, 225–242.

Griffin, J.E. & R.C. Oomen (2011) Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics* 160, 58–68.

Hayashi, T. & N. Yoshida (2005) On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.

Kristensen, D. (2010) Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* 26, 60–93.

Palandri, A. (2009) Sequential conditional correlation: Inference and evaluation. *Journal of Econometrics* 153, 122–132.

Reiss, M. (2011) Asymptotic equivalence for inference on the volatility from noisy observations. *Annals of Statistics* 39, 772–802.

Tao, M., Y. Wang, Q. Yao, & J. Zou (2011) Large volatility matrix inference via combining low-frequency and high-frequency approaches. *Journal of the American Statistical Association* 106, 1025–1040.

Wang, Y. & J. Zou (2010) Vast volatility matrix estimation for high-frequency financial data. *Annals of Statistics* 38, 943–978.

Zhang, L. (2006) Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12, 1019–1043.

Zhang, L. (2011) Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics* 160, 33–47.

Zhang, L., P.A. Mykland, & Y. Aït-Sahalia (2005) A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.

https://doi.org/10.1017/S0266466612000746 Published online by Cambridge University Press