

# Large Volatility Matrix Estimation with Factor-Based Diffusion Model for High-Frequency Financial data

DONGGYU KIM<sup>\*</sup>, YI LIU<sup>\*\*</sup>, and YAZHEN WANG<sup>†</sup>

*College of Business, Korea Advanced Institute of Science and Technology (KAIST), Seoul, Korea. E-mail: <sup>\*</sup>dkim0329@gmail.com*

*Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706. E-mail: <sup>\*\*</sup>clarapku@gmail.com; <sup>†</sup>yzwang@stat.wisc.edu*

Large volatility matrices are involved in many finance practices, and estimating large volatility matrices based on high-frequency financial data encounters the “curse of dimensionality”. It is a common approach to impose a sparsity assumption on the large volatility matrices to produce consistent volatility matrix estimators. However, due to the existence of common factors, assets are highly correlated with each other, and it is not reasonable to assume the volatility matrices are sparse in financial applications. This paper incorporates factor influence in the asset pricing model and investigates large volatility matrix estimation under the factor price model together with some sparsity assumption. We propose to model asset prices by assuming that asset prices are governed by common factors and that the assets with similar characteristics share the same association with the factors. We then impose some reasonable sparsity condition on the part of the volatility matrices after accounting for the factor contribution. Under the proposed factor-based model and sparsity assumption, we develop an estimation scheme called “blocking and regularizing”. Asymptotic properties of the proposed estimator are studied, and its finite sample performance is tested via extensive numerical studies to support theoretical results.

*Keywords:* Multiple-scale realized volatility, kernel realized volatility, pre-averaging realized volatility, factor model, regularization, sparsity, adaptive threshold, diffusion, integrated volatility.

## 1. Introduction

High-frequency finance provides academic researchers and industry practitioners with new resources and tools to study financial markets. With high-frequency financial data we can better model the volatility dynamics and efficiently estimate volatilities. There is a large volume of literature on volatility estimation based on high-frequency data. For a single asset, the estimators for an integrated volatility include realized volatility (Anderson *et al.* (2003) and Barndorff-Neilsen and Shepard (2002)), bi-power realized variation (Barndorff-Neilsen and Shepard (2006)), two-time scale realized volatility (Zhang, Mykland and Ait-Sahalia (2005)), multiple-time scale realized volatility (Zhang (2006)),

wavelet realized volatility (Fan and Wang (2007)), kernel realized volatility (Barndorff-Neilsen *et al.* (2008)), Fourier realized volatility (Mancino and Sanfelici (2008)), pre-averaging realized volatility (Jacod *et al.* (2009)), and quasi maximum likelihood estimator (Xiu (2010)). For multiple assets, methods for estimating an integrated co-volatility consist of Hayashi-Yoshida estimator (Hayashi and Yoshida (2005)), a quasi maximum likelihood estimator (Ait-Sahalia, Fan, and Xiu (2010)), multi-scale realized co-volatility based on previous tick data synchronization (Zhang (2011)), preaveraging realized volatility (Christensen, Kinnebrock, and Podolskij (2010)), preaveraging Hayashi-Yoshida estimator (Christensen, Podolskij, and Vetter (2013)), Fourier covariance estimator (Mancino and Sanfelici (2011)), realized kernel volatility estimator based on refresh time scheme (Barndorff-Neilsen *et al.* (2011)), and local method of moments (Bibinger *et al.* (2014)).

In financial markets, a large number of assets are often involved in asset pricing, portfolio allocation, and risk management, where the estimation of their volatility matrix plays a pivotal role in financial applications. Estimators developed for small volatility matrices often become inconsistent when the number of assets is comparable to or even exceed the number of observations (Wang and Zou (2010)). This is the so-called “curse of dimensionality”. Wang and Zou (2010) proposed the averaged realized volatility matrix (ARVM) with good convergence rates using thresholding estimation procedures. In Tao *et al.* (2013a, 2013b), the multiple-time scale realized volatility matrix (MSRVM) together with thresholding rules was proposed to achieve the optimal convergence rate. Kim *et al.* (2016) introduced the thresholding rules for kernel realized volatility matrix (KRVM) estimator and pre-averaging volatility matrix (PARVM) estimator and established the asymptotic theory for these thresholded estimators. These existing large volatility matrix estimators, such MSRVM, PARVM and KRVM, were used to derive consistent estimators by regularizing them under the assumption that the true volatility matrix is sparse. Although the sparsity assumption and regularizing technique have been widely employed to overcome the “curse of dimensionality” (see, for example, Bickel and Levina (2008a, b), Cai and Liu (2011), Cai and Zhou (2012), Kim *et al.* (2015, 2016), Tao *et al.* (2013a, b), and Wang and Zou (2010)), it may not be reasonable to simply impose the usual sparsity assumption on financial volatility matrices, due to the existence of common factors and high correlation among assets.

Factor models have been widely used in both theoretical and empirical finance. For example, using arbitrage pricing theory, Ross (1976, 1977) proposed a multi-factor model to show that the excessive return of any asset is a linear function of various macro economic factors. Since then there are many papers contributed to the construction of factors, see, for example, Engle and Watson (1981), Chamberlain (1983), Chamberlain and Rothschild (1983), Diebold and Nerlove (1989), Fama and French (1992, 1993), Aguilar and West (2000), and Stock and Watson (2005). Fan *et al.* (2008, 2013) proposed to use factor models for estimating high-dimensional covariance matrices and showed that utilization of factor models improves the performance of the covariance matrix estimator. Despite the rich development about factor models and the recent progress on large volatility matrix estimation, it is remarkable that there is little research on incorporating factors into volatility estimation with high-frequency financial data. This motivates us to study large volatility matrix estimation with factor models for high-

frequency financial data.

Intuitively the expected return of an asset relates to economic factors such as sector and industry classification, firm size, price to book ratios, etc. (see among others, Fama and French (1992)), it is thus expected that assets with similar characteristics tend to have similar expected returns. Therefore, we propose a factor-based diffusion process to account for the grouping effect in asset pricing dynamics, and then impose a sparsity condition on the part of the volatility matrices after accounting for the factor contribution. Based on the structure knowledge of the volatility matrices, we develop a “blocking and regularizing” strategy to construct a volatility matrix estimator. The proposed estimator is shown to have better performance than existing ones in presence of common factors.

The rest of the paper is organized as following. In Section 2, a factor-based diffusion process is proposed to model the asset pricing dynamics. Section 3 discusses in detail how blocking and regularizing are used to construct a volatility matrix estimator. Section 4 investigates the asymptotic properties of the proposed estimator. In Section 5, an extensive simulation study is conducted to check the finite sample performance of the proposed estimator. Section 6 illustrates large volatility matrix estimation for high-frequency data on 200 stocks. Section 7 features the conclusion. Section 8 collects all technical proofs.

## 2. The model set-up

Consider  $B$  different groups with  $p_b$  ( $b = 1, \dots, B$ ) assets in each group. In the  $b$ -th group, let  $X_{b_i}(t)$ ,  $i = 1, \dots, p_b$ ,  $b_i = \sum_{k=0}^{b-1} p_k + i$ , be the true log price at time  $t$  for the  $b_i$ -th asset, and denote by  $\mathbf{X}_b(t) = (X_{b_1}(t), \dots, X_{b_{p_b}}(t))^T$  the vector of true log prices. Let  $p$  be the total number of assets, that is,  $p = \sum_{b=1}^B p_b$ . Let  $\mathbf{X}^T(t) = (\mathbf{X}_1^T(t), \dots, \mathbf{X}_B^T(t))$  be the true log prices of the  $p$  assets. Common approach in finance assumes that the log stock price follows a continuous-time diffusion model with a drift term. In order to account for grouping factors, we consider the following model

$$d\mathbf{X}(t) = \boldsymbol{\mu}(t)dt + \boldsymbol{\vartheta}^T(t)d\mathbf{W}_t^* + \boldsymbol{\sigma}^T(t)d\mathbf{W}_t, \quad (2.1)$$

where  $\boldsymbol{\mu}(t) = (\mu_1(t), \dots, \mu_p(t))^T$  is a drift vector,  $\boldsymbol{\sigma}(t)$  is a  $p$  by  $p$  matrix,  $\boldsymbol{\vartheta}(t)$  is a  $B$  by  $p$  matrix,  $\mathbf{W}_t^*$  and  $\mathbf{W}_t$  are  $B$ -dimensional Brownian motion and  $p$ -dimensional Brownian motion, respectively, and they are independent. The spot volatility matrix of  $\mathbf{X}(t)$  is

$$\boldsymbol{\gamma}(t) = (\gamma_{ij}(t))_{1 \leq i, j \leq p} = \boldsymbol{\sigma}^T(t)\boldsymbol{\sigma}(t) + \boldsymbol{\vartheta}^T(t)\boldsymbol{\vartheta}(t),$$

while the integrated volatility matrix, or quadratic variation of  $\mathbf{X}(t)$ , is

$$\begin{aligned} [\mathbf{X}, \mathbf{X}]_t &= \int_0^t \boldsymbol{\gamma}(s)ds = \left( \int_0^t \gamma_{ij}(s)ds \right)_{1 \leq i, j \leq p} \\ &= \left( \int_0^t \sigma_{ij}(s)ds \right)_{1 \leq i, j \leq p} + \left( \int_0^t \vartheta_{ij}(s)ds \right)_{1 \leq i, j \leq p} \\ &= \int_0^t \boldsymbol{\sigma}^T(s)\boldsymbol{\sigma}(s)ds + \int_0^t \boldsymbol{\vartheta}^T(s)\boldsymbol{\vartheta}(s)ds. \end{aligned}$$

The parameter of interest is the integrated volatility matrix over time  $[0, 1]$ ,

$$\mathbf{\Gamma} = \int_0^1 \boldsymbol{\sigma}^T(t)\boldsymbol{\sigma}(t)dt + \int_0^1 \boldsymbol{\vartheta}^T(t)\boldsymbol{\vartheta}(t)dt = \mathbf{\Sigma} + \mathbf{\Theta}. \quad (2.2)$$

$\mathbf{\Theta}$  defined in (2.2) accounts for the factor influence on the volatility matrix and has the following blocking structure. For  $a, b = 1, \dots, B$ ,

$$\mathbf{\Theta} = (\Theta_{ij})_{1 \leq i, j \leq p}, \quad \Theta_{ij} = \theta_{ab}, \quad \forall i = a_1, \dots, a_{p_a}, j = b_1, \dots, b_{p_b}. \quad (2.3)$$

$\theta_{bb}$ 's capture the factor influence on all the assets in the  $b$ -th group, while  $\theta_{ab}$ ,  $a \neq b$ , account for the cross-sectional influence between the  $a$ -th and  $b$ -th groups. On the other hand,  $\mathbf{\Sigma}$  consists of volatilities after accounting for the factor contribution, which cannot be attributed to the factors.

**Remark 1.** Recently, Ait-Sahalia and Xiu (2015) and Fan *et al.* (2016) introduced some factor model based on the high-frequency observations, which has the low-rank plus sparse structure. The sparse structure can be represented by the block diagonal matrix. That is, the integrated volatility matrix from their factor model consists of the general factor structure and the block diagonal sparse structure. So both the proposed volatility structure in this paper and their volatility structure have a few common factors which account for the highly correlated large volatility matrix, and their remaining risks are explained by the sparse matrix. On the other hand, in our proposed factor model, the low-rank volatility matrix  $\mathbf{\Theta}$  has some block structure which can be considered as the sub-class of the low-rank matrix in Ait-Sahalia and Xiu (2015) and Fan *et al.* (2016). For example, choosing the factor loading matrix as the membership matrix such as the factor loading matrix in (5.1), their low-rank matrix will become the factor volatility matrix  $\mathbf{\Theta}$  in (2.3). That is, the assets in the same group have the same latent factor process. With this additional structure, we investigate how to better accommodate the sparse structure of large volatility matrices.

**Remark 2.** In this paper, we assume that the group membership is known. In practice, it becomes a huge issue to identify the group membership, and the performance of the proposed estimator depends on the choice of the group membership. In the finance market, the assets in the same sector tend to have some common behavior. For example, the companies in the financial sector obey the same financial regulations and face similar market risks. In light of this, we propose to use the global industry classification standard (GICS) sector as the group membership in this paper. According to the empirical study, the GICS sector can explain the grouping effect well. In addition, our model has some robustness to the choice of the group membership. For example, if we can identify the sub-groups whose members have the common factor, the proposed estimators can enjoy the theoretical results developed in Section 4.2. Furthermore, the numerical study supports the claim.

The real-time high-frequency trading prices have two features. First, instead of observing the true price  $X_i(t)$ ,  $i = 1, \dots, p$ , we observe  $Y_i(t_{i,h})$ , its noisy version at discrete

times  $t_{i,h}, h = 1, \dots, n_i$ . Second, different assets are traded at different times, that is,  $t_{i,h} \neq t_{j,h}$  for any  $i \neq j$ . This is the so-called non-synchronization problem. In light of these, we assume that the observed price data  $Y_i(t_{i,h})$  obey the following model

$$Y_i(t_{i,h}) = X_i(t_{i,h}) + \epsilon_i(t_{i,h}) \quad \text{for } i = 1, \dots, p, h = 1, \dots, n_i, \quad (2.4)$$

where  $\epsilon_i(\cdot)$  are noises with mean zero and variance  $\eta_i$ .

### 3. Blocked and regularized volatility matrix estimator

Under the proposed price model, we construct a volatility matrix estimator in three stages. First, a realized volatility matrix is constructed from the noisy high-frequency data. The aforementioned challenges, non-synchronicity of data and noise contamination, are handled in this stage. There are several well-performing realized volatility matrix estimators, for example, multi-scale realized co-volatility based on previous tick data synchronization (Zhang (2011)), the quasi maximum likelihood estimator (Ait-Sahalia, Fan, and Xiu (2010)), pre-averaging realized volatility (Christensen, Kinnebrock, and Podolskij (2010)), realized kernel volatility estimator based on refresh time scheme (Barndorff-Neilsen *et al.* (2011)), and preaveraging Hayashi-Yoshida estimator (Christensen, Kinnebrock, and Podolskij (2010) and Christensen, Podolskij, and Vetter (2013)). In this paper, we take one of these estimators as the input realized volatility matrix estimator  $\widehat{\Gamma}$ . Second, we decompose the realized volatility matrix to separate the part account for the factor contribution and the left-over part, which are referred as the factor volatility matrix and non-factor volatility matrix, respectively. In the third stage, we regularize the non-factor volatility matrix. The second and third stages are discussed in detail in Section 3.1 below.

#### 3.1. Blocking and regularizing

First, we decompose a chosen realized volatility matrix into the factor and non-factor contributions. Then we impose a sparsity condition on the non-factor volatility matrix and threshold its corresponding estimator. We call the constructed estimator *blocked and regularized volatility matrix* (BRVM) estimator.

##### 3.1.1. Decompose $\widehat{\Gamma}$ : blocking

Taking advantage of the block structure of the factor volatility matrix described by (2.3), we estimate the cross-sectional volatility between groups  $a$  and  $b$  by averaging corresponding elements in the realized volatility matrix estimator  $\widehat{\Gamma} = (\widehat{\Gamma}_{ij})_{i,j=1,\dots,p}$  as follows,

$$\widehat{\theta}_{ab} = \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} \widehat{\Gamma}_{ij}, \quad a, b = 1, \dots, B.$$

Let  $\widehat{\Theta}$  be the estimator of  $\Theta$ , specifically,

$$\widehat{\Theta} = \left( \widehat{\Theta}_{ij} \right)_{1 \leq i, j \leq p} \quad \text{and} \quad \widehat{\Theta}_{ij} = \widehat{\theta}_{ab}, \quad \forall i = a_1, \dots, a_{p_a}, j = b_1, \dots, b_{p_b}.$$

The non-factor volatility matrix  $\Sigma$  can then be estimated by

$$\widehat{\Sigma} = \widehat{\Gamma} - \widehat{\Theta}.$$

Equivalently, we have decomposed the realized volatility matrix estimator as

$$\widehat{\Gamma} = \widehat{\Theta} + \widehat{\Sigma}. \quad (3.1)$$

### 3.1.2. Sparsity condition and regularizing

Imposing the sparsity condition and thresholding realized volatility matrices are a common approach to obtain consistent estimators of large matrices (Bickel and Levina (2008a, b), Cai and Liu (2011), Cai and Zhou (2012), Kim *et al.* (2015, 2016), Tao *et al.* (2013a, b), and Wang and Zou (2010)). In finance practices, imposing the sparsity condition on the entire volatility matrix implies that only a small number of assets are correlated with each other, which is not realistic. Instead of assuming the sparsity for the entire volatility matrix, we impose the sparsity condition on the non-factor volatility matrix, that is, the part of the volatility matrix after accounting for the factor contribution. Specifically, we assume that  $\Sigma$  satisfies the following condition

$$\max_{a_1 \leq i \leq a_{p_a}} \sum_{j=b_1}^{b_{p_b}} |\Sigma_{ij}|^\delta |\Sigma_{ii} \Sigma_{jj}|^{(1-\delta)/2} \leq \Psi \frac{\pi(p)}{B}, \quad \forall a, b = 1, \dots, B, \quad (3.2)$$

where  $0 \leq \delta < 1$ ,  $\Psi$  is a positive random variable with  $\mathbb{E}|\Psi|^\beta \leq C$  for some constant  $C$  and  $\beta \geq 2$  ( $\beta$  will be given in Section 4), and  $\pi(p)$  is a deterministic function that grows very slowly in  $p$ . Examples of  $\pi(p)$  include 1,  $\log(p)$  and a very small power of  $p$ .

**Remark 3.** The condition (3.2) is imposed on each block of  $\Sigma$ . Summing (3.2) over all the blocks, we have

$$\max_{1 \leq i \leq p} \sum_{j=1}^p |\Sigma_{ij}|^\delta |\Sigma_{ii} \Sigma_{jj}|^{(1-\delta)/2} \leq \Psi \pi(p). \quad (3.3)$$

This condition implies the  $\ell_1$  bound as follows:

$$\max_{1 \leq i \leq p} \sum_{j=1}^p |\Sigma_{ij}| \leq \Psi \pi(p).$$

If  $\Sigma$  satisfies the condition (3.2), its important elements are those who exceed certain threshold. Therefore, we regularize  $\widehat{\Sigma}$  by retaining its elements whose absolute values

exceed a given threshold value and replacing the others by zero. Specifically, we define the thresholded  $\widehat{\Sigma}$  by

$$\begin{aligned}\widetilde{\Sigma} &= \mathcal{T}_{\varpi}^H[\widehat{\Sigma}] = \left( \widehat{\Sigma}_{ij} \mathbf{1}(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}) \right)_{1 \leq i, j \leq p} \quad \text{or} \\ \widetilde{\Sigma} &= \mathcal{T}_{\varpi}^S[\widehat{\Sigma}] = \left( \left( \widehat{\Sigma}_{ij} - \text{sign}(\widehat{\Sigma}_{ij}) \varpi_{ij} \right) \mathbf{1}(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}) \right)_{1 \leq i, j \leq p},\end{aligned}\quad (3.4)$$

where  $\mathbf{1}(\cdot)$  is an indicator function,  $\varpi_{ij} = \lambda \sqrt{(\widehat{\Sigma}_{ii} \vee 0)(\widehat{\Sigma}_{jj} \vee 0)}$  is adaptive thresholding parameter, and  $\lambda$  will be given in Section 4. The two estimators are called hard and soft thresholding rules, respectively, and the adaptive thresholding rules are based on the correlation structure. Both thresholding rules retain some significant values who exceed certain thresholding level. For example, the  $(i, j)$ th element of the hard thresholding estimator is equal to  $\widehat{\Sigma}_{ij}$  if it is greater than the threshold and zero otherwise. On the other hand, the soft thresholding rule utilizes a smooth function, and so its  $(i, j)$ th element is  $\widehat{\Sigma}_{ij} - \text{sign}(\widehat{\Sigma}_{ij}) \varpi_{ij}$  if it is greater than the threshold.

Replacing  $\widehat{\Sigma}$  in (3.1) with its thresholded version  $\widetilde{\Sigma}$ , we obtain the *blocked and regularized volatility matrix* (BRVM) estimator

$$\widetilde{\Gamma} = \widehat{\Theta} + \widetilde{\Sigma}. \quad (3.5)$$

The asymptotic behavior of the BRVM estimator  $\widetilde{\Gamma}$  depends on the underlying realized volatility matrix estimator  $\widehat{\Gamma}$  and will be investigated in Section 4.

## 4. Asymptotic theorems

### 4.1. Technical conditions

For any  $U = (U_{ij})_{i, j=1, \dots, p} \in \mathbb{R}^{p \times p}$ , denote by  $\|U\|_F = \sqrt{\text{tr}(U^T U)}$  its Frobenius norm, and define the weighted quadratic norm (Fan *et al.* (2008, 2013)),

$$\|U\|_{\Gamma}^2 = p^{-1} \|\Gamma^{-1/2} U \Gamma^{-1/2}\|_F^2.$$

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

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

**Assumption 1.** For some  $\beta \geq 2$ ,

$$\begin{aligned}\max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} \mathbb{E} [|\vartheta_{ii}(t)|^\beta] &< \infty, & \max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} \mathbb{E} [|\sigma_{ii}(t)|^\beta] &< \infty, \\ \max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} \mathbb{E} [|\mu_i(t)|^{2\beta}] &< \infty, & \max_{1 \leq i \leq p} \mathbb{E} [|\epsilon_i(t_{i,h})|^{2\beta}] &< \infty.\end{aligned}\quad (4.1)$$

**Assumption 2.** Assume that for some fixed constant  $C_\Sigma > 0$ ,

$$\min_{1 \leq i \leq p} \Sigma_{ii} \geq C_\Sigma.$$

**Assumption 3.** There exist positive constants  $C_1$  and  $C_2$  such that

$$C_1 \leq \frac{p_a}{p_b} \leq C_2, \quad \forall a, b = 1, \dots, B.$$

**Remark 4.** Assumption 1 imposes moment conditions on the price process and noises in order to establish the asymptotic theory. In particular, the condition (4.1) implies

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

for some  $\beta \geq 2$ .

**Remark 5.** Assumption 3 allows the number of assets in each group to vary, as constants  $C_1$  and  $C_2$  may be substantially different.

## 4.2. Asymptotic theorems

To investigate the asymptotic behavior of the BRVM estimator, we need to establish the moment bound for its entries. Theorem 4.1 provides the element-wise bound for estimators of  $\Theta$  and  $\Sigma$ .

**Theorem 4.1.** Under the models (2.1) and (2.4), suppose that the moment bound

$$\max_{1 \leq i, j \leq p} \mathbb{E} \left[ |\widehat{\Gamma}_{ij} - \Gamma_{ij}|^\beta | n \right] \leq C n^{-\beta/4} \text{ a.s.}, \quad (4.2)$$

condition (3.2), and Assumptions 1-3 are met. Then we have the following element-wise moment bound for  $\widehat{\Theta}$  and  $\widehat{\Sigma}$

$$\mathbb{E} \left[ |\widehat{\Theta}_{ij} - \Theta_{ij}|^\beta | n \right] \leq e_n^\beta, \quad \mathbb{E} \left[ |\widehat{\Sigma}_{ij} - \Sigma_{ij}|^\beta | n \right] \leq e_n^\beta \text{ a.s.},$$

where  $e_n = C [n^{-1/4} + \pi(p)/p]$  for some constant  $C$  free of  $n$  and  $p$ .

**Remark 6.** Theorem 4.1 shows that above asymptotic results hold for any volatility matrix estimators satisfying the moment bound condition (4.2). For example, the pre-averaging realized volatility matrix (PARVM) (Christensen, Kinnebrock, and Podolskij (2010)), multi-scale realized volatility matrix (MSRVM) (Zhang (2011)), and realized kernel volatility matrix (KRVM) (Barndorff-Neilsen *et al.* (2011)) estimators satisfy the moment bound condition (4.2) with some synchronization scheme such as the generalized sampling time (Ait-Sahalia *et al.* (2010)) under Assumption 1 (see Theorems 1 and 3 (Kim *et al.* (2016))). Also the pre-averaging Hayashi-Yoshida estimator (Christensen, Kinnebrock, and Podolskij (2010) and Christensen, Podolskij, and Vetter (2013)) can meet the moment bound condition (4.2) (see Proposition 1 in Christensen, Kinnebrock, and Podolskij (2010)), without any synchronization scheme requirement.



As Fan *et al.* (2013) discussed, it is possible to estimate the large integrated volatility matrix in terms of the weighted quadratic norm. This paper considers the weighted quadratic norm  $\|\cdot\|_{\mathbf{\Gamma}}$ . We need the following technical assumptions.

**Assumption 4.** *There is some constant  $C > 0$  free of  $n$  and  $p$  such that*

$$\lambda_p(\mathbf{\Gamma}) > C,$$

where  $\lambda_j(\mathbf{\Gamma})$  is the  $j$ -th largest eigenvalue for  $\mathbf{\Gamma}$ .

**Assumption 5.** *There are some positive constants  $c_1, c_2 > 1$  free of  $n$  and  $p$  such that*

$$1/c_1 \leq p^{-2} \|\Theta\|_F^2 \leq c_1$$

and

$$1/c_2 < \lambda_1(\Theta)/\lambda_B(\Theta) < c_2.$$

**Remark 7.** Assumption 4 indicates that the integrated volatility matrix  $\mathbf{\Gamma}$  is positive definite which is a usual condition in analyzing volatility matrices. Assumption 5 implies that the eigenvalues of the factor volatility matrix  $\Theta$  diverge with  $p/B$  order, which is called the pervasive condition. The pervasive condition is usually imposed on analyzing the approximate factor models (see Fan *et al.* (2013)). Furthermore, we consider the factor model to account for the fact that the assets are highly correlated and so the volatility matrix is not sparse. Thus, the factor volatility matrix  $\Theta$  is a non-sparse matrix satisfying Assumption 5. In light of these, Assumptions 4 and 5 are not restrictive.

The theorem below derives the convergence rate for  $\mathbf{\Gamma}$ .

**Theorem 4.2.** *Assume that data are observed from the models (2.1) and (2.4), conditions (3.2) and (4.1), and Assumptions 1-5 are met, and  $p^{4/\beta}e_n = o(1)$ . Then we have the following convergence rate for  $\tilde{\mathbf{\Gamma}}$  in (3.5)*

$$\|\tilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\mathbf{\Gamma}} = O_p \left( [B^{1/2}p^{1/2}e_n^2 + B^{1/4}e_n][1 + B^{3/2}e_n] + \pi(p)\lambda^{1-\delta} \right),$$

where  $e_n$  is given in Theorem 4.1, and the thresholding parameter  $\lambda$  is of order  $p^{2/\beta}e_n$ .

**Remark 8.** Theorem 4.2 shows that the BRVM estimator has the convergence rate  $[B^{1/2}p^{1/2}e_n^2 + B^{1/4}e_n][1 + B^{3/2}e_n] + \pi(p)\lambda^{1-\delta}$  in terms of the weighted quadratic norm. To obtain this rate, we do not need any condition for the number of factors,  $B$ , but to obtain the consistency, we may need some condition on  $B$  and  $p$ . In the stock market, we can observe that the assets are governed by a few common factors such as sector and industry classification, firm size, and price to book ratios. So the number,  $B$ , of factors is much less than the number of stocks, and so it is reasonable to assume  $B = O(1)$ . Then the convergence rate in Theorem 4.2 becomes  $p^{1/2}e_n^2 + \pi(p)\lambda^{1-\delta}$ . Thus, Theorem 4.2

shows that the BRVM estimator can still be consistent so long as  $p = o(n)$ . On the other hand, in Fan *et al.* (2013), the principal orthogonal complement thresholding estimator can be consistent as long as  $p = o(n^2)$ . The difference is due to the contamination of the microstructure noise in the observed data. In fact, if the true log prices  $X_i$ 's are observed, the BRVM estimator only requires the same condition  $p = o(n^2)$  to obtain the consistency.

## 5. A simulation study

We conducted a simulation study to check the finite sample performance of the BRVM estimators constructed using multi-scale realized volatility matrix (MSRVM) (Zhang (2011)), pre-averaging realized volatility matrix (PARVM) (Christensen, Kinnebrock, and Podolskij (2010)), realized kernel volatility matrix (KRV) (Barndorff-Nielsen *et al.* (2011)), and preaveraging Hayashi-Yoshida volatility matrix (PHYVM) (Christensen, Podolskij, and Vetter (2013)). The simulation study adopts the model set-up in Wang and Zou (2010). Section 5.1 describes the simulation model and procedures, and Section 5.2 summarizes the simulation results.

### 5.1. The simulation model

Since the drift term is of negligible order in integrated volatility estimation, so we set  $\boldsymbol{\mu}(t) = 0$  in our simulation study, and assume the true log prices  $\mathbf{X}(t)$  of  $p$  assets to follow the model,

$$\mathbf{X}(t) = \int_0^t \boldsymbol{\vartheta}^T(s) d\mathbf{W}_s^* + \int_0^t \boldsymbol{\sigma}^T(s) d\mathbf{W}_s,$$

where  $\mathbf{W}_t = (W_{1t}, \dots, W_{pt})^T$  is a  $p$ -dimension standard Brownian motion, and  $\mathbf{W}_t^*$  is a  $B$ -dimensional Brownian motion that is independent of  $\mathbf{W}_t$ . We take  $\boldsymbol{\sigma}(t)$  as Cholesky decomposition of  $\boldsymbol{\varsigma}(t) = (\varsigma_{ij}(t))_{i,j=1,\dots,p}$  as below. First, the diagonal elements of  $\boldsymbol{\varsigma}(t)$  are generated from geometric Ornstein-Uhlenbeck processes, the volatility process in Nelson's GARCH diffusion limit model (Wang (2002)), the sum of two CIR processes (Cox, Ingersoll and Ross (1985) and Barndorff-Nielsen and Shepard (2002)), and two-factor log-linear stochastic volatility process (Huang and Tauhen (2005)) with leverage effect. Then we define off-diagonal elements of  $\boldsymbol{\varsigma}(t)$  by

$$\varsigma_{ij}(t) = \{\kappa(t)\}^{|i-j|} \sqrt{\varsigma_{ii}(t)\varsigma_{jj}(t)}, \quad 1 \leq i \neq j \leq p,$$

where  $\kappa(t)$  is given by

$$\begin{aligned} \kappa(t) &= \frac{e^{2u(t)} - 1}{e^{2u(t)} + 1}, \quad du(t) = 0.03[0.64 - u(t)]dt + 0.118u(t)dW_{\kappa,t}, \\ W_{\kappa,t} &= \sqrt{0.96}W_{\kappa,t}^0 - 0.2 \sum_{i=1}^p W_{it}/\sqrt{p}; \end{aligned}$$

$W_{\kappa,t}^0$  is a 1-dimensional Brownian motion independent of  $\mathbf{W}_t$  and  $\mathbf{W}_t^*$ . See Wang and Zou (2010) for details.

Define the factor volatility process  $\vartheta(t)$  as

$$\vartheta(t) = \vartheta^f(t)\mathbf{L}^T \in \mathbb{R}^{B \times p},$$

where

$$\mathbf{L} = \begin{pmatrix} \mathbf{J}_{p/B} & 0 & \cdots & 0 \\ 0 & \mathbf{J}_{p/B} & & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \mathbf{J}_{p/B} \end{pmatrix} \in \mathbb{R}^{p \times B}, \quad \mathbf{J}_{p/B} = (1, \dots, 1)^T \in \mathbb{R}^{p/B}, \quad (5.1)$$

and  $\vartheta^f(t) \in \mathbb{R}^{B \times B}$  is generated similar to  $\sigma(s)$ . Specifically,  $\vartheta^f(t)$  is Cholesky decomposition of  $(\varsigma_{ij}^f(t))_{i,j=1,\dots,B}$ , and the diagonal elements of  $(\varsigma_{ij}^f(t))_{i,j=1,\dots,B}$  were generated from the four processes (geometric Ornstein-Uhlenbeck processes, the sum of two CIR processes, the volatility process in Nelson's GARCH diffusion limit model, and two-factor log-linear stochastic volatility process). Then we generated the off-diagonal elements as follows:

$$\varsigma_{ij}^f(t) = \kappa^f(t) \sqrt{\varsigma_{ii}^f(t) \varsigma_{jj}^f(t)}, \quad 1 \leq i \neq j \leq B,$$

where  $\kappa^f(t)$  is defined

$$\begin{aligned} \kappa^f(t) &= \frac{e^{2u^f(t)} - 1}{e^{2u^f(t)} + 1}, \quad du^f(t) = 0.03[0.64 - u^f(t)]dt + 0.118u^f(t)dW_{\kappa,t}^f, \\ W_{\kappa,t}^f &= \sqrt{0.96}W_{\kappa,t}^{f0} - 0.2 \sum_{i=1}^B W_{it}^*/\sqrt{p}; \end{aligned}$$

$W_{\kappa,t}^{f0}$  is a 1-dimensional Brownian motion independent of  $W_{\kappa,t}^0$ ,  $\mathbf{W}_t$ , and  $\mathbf{W}_t^*$ .

Finally, the high-frequency data  $Y_i(t_{i,h})$  are simulated from the model (2.4) with noise  $\epsilon_i(t_{i,h})$  drawing from independent normal distributions with mean zero and standard deviation  $\eta_i$  which is given by

$$\eta_i = a \sqrt{\int_0^1 (\vartheta_{ii}(t) + \varsigma_{ii}(t)) dt}, \quad i = b_1, \dots, b_{p_b}, b = 1, \dots, B,$$

where  $a$  is the relative noise level ranges from 0 to 0.1 in the simulation study. We generated non-synchronized observations as follows. First,  $p$  random proportions  $m_i$ ,  $i = 1, \dots, p$ , are generated independently from a uniform distribution on  $(0.5, 1)$ . Second, for the  $i$ -th asset, we selected  $n_i = \lfloor m_i n^{all} \rfloor$  observations from the synchronized sample,  $\{Y_i(t_r), r = 1, \dots, n^{all}\}$ , randomly and denote the selected observations by  $\{Y_i(t_{i,r}), r = 1, \dots, n_i, i = 1, \dots, p\}$  which are non-synchronized noisy data.

## 5.2. Simulation results

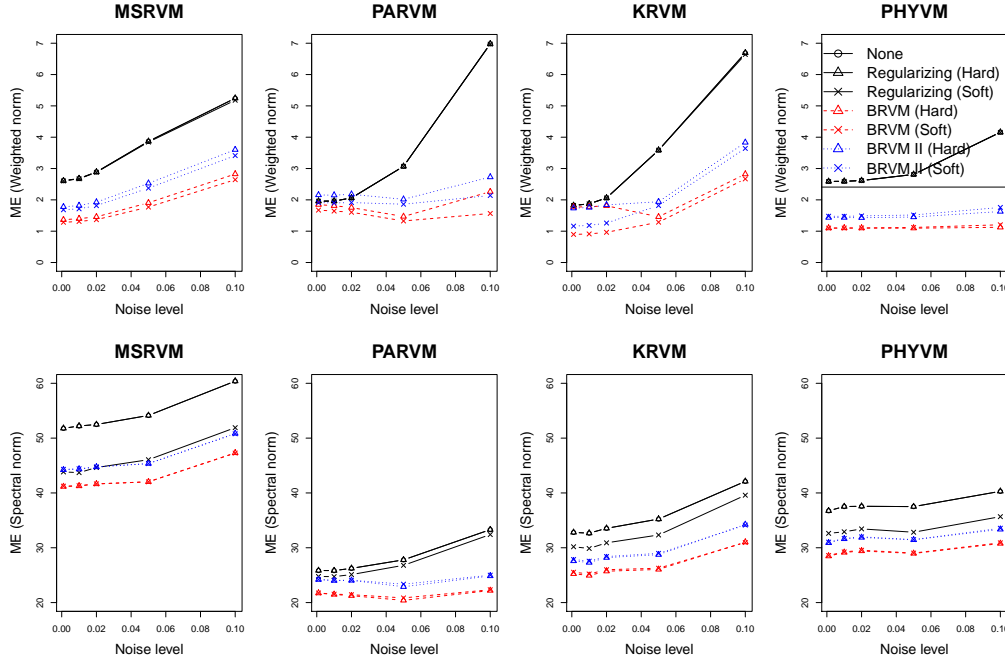
In the simulation study, we chose  $p = 256$  and  $B = 16$ , fixed  $n^{all} = 2000$ , and selected noise levels to be 0.001, 0.01, 0.02, 0.05, and 0.1.

We employed the simulated data  $Y_i(t_{i,h})$  to compute the BRVM estimator defined in Section 3 by using MSRVM, PARVM, and KRVM defined in Kim *et al.* (2016) and PHYVM in Christensen, Podolskij, and Vetter (2013). To synchronize the simulated data, the refresh time scheme (Barndorff-Neilsen *et al.* (2011)) was employed, and we found the average number of synchronized data points over the 256 simulated assets was equal to 308. To calculate MSRVM, PARVM, and KRVM, we used synchronized observations. We used the weight function  $g(x) = x \wedge (1 - x)$  for the PARVM and PHYVM estimators and the Parzen kernel,

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & \text{if } 0 \leq x \leq 1/2 \\ 2(1 - x)^3 & \text{if } 1/2 \leq x \leq 1 \\ 0 & \text{if } x > 1, \end{cases}$$

for the KRVM estimator. After realized volatility matrix estimators were constructed, we took different blocking and regularization strategies: 1) none; 2) regularizing only; 3) blocking and regularizing with the correct number of groups (BRVM); and 4) blocking and regularizing with twice as many groups as the true grouping structure (BRVM II). The whole simulation procedure was repeated 500 times, and the  $\ell_2$ -norms (spectral norms) and the weighted quadratic norms,  $\|\cdot\|_{\Gamma}$ , of the difference between estimator and true volatility matrix were computed and averaged over the 500 repetitions. We selected the tuning parameters for MSRVM, PARVM, KRVM, and PHYVM and the threshold level  $\lambda$  by minimizing the corresponding mean errors (ME).

Figure 1 plots the average  $\ell_2$  (spectral norm) errors and weighted quadratic errors against noise levels for thresholded estimators constructed using MSRVM, PARVM, KRVM and PHYVM, and compares the performance of these estimators under different blocking and regularizing strategies, and Table 1 reports their numerical results. We found that blocking and regularizing scheme significantly improves volatility matrix estimation. In particular, the largest performance improvement is obtained when the grouping structure is correctly identified. Moreover, even when assets are divided into twice as many groups as the true grouping structure, the BRVM II estimators still have good performances. In practice, when the true grouping structure of the assets is unknown, dividing assets into smaller groups helps maintain the homogeneity of volatility matrix within each block and improves the volatility estimation. On the other hand, regularizing only does not improve much the performance of the estimators in terms of ME. This may be due to the fact that the volatility matrix is no longer sparse with the presence of factor influence. These numerical results are in consistent with our theoretical findings. Regarding choosing realized volatility matrix estimators, we found that the pre-averaging type estimators (PARVM and PHYVM) have better performance, while the PHYVM has the smallest weighted quadratic norm error.



**Figure 1.** The ME plots against five relative noise levels for MSRVM, KRVM, PARVM, and PHYVM estimators, and their corresponding thresholded estimators and BRVM estimators, with  $p = 256$ , and  $n^{all} = 2000$ .

## 6. An empirical study

In this section, we applied the BRVM estimators to high-frequency trading data observed on 200 assets over a period of 63 days in 2013. We selected top 200 large trading volume stocks among S&P 500, and used the previous-tick times to compute 1-min log returns (Wang and Zou (2010)). The data was taken from the Wharton Data Service (WRDS) system. To evaluate the BRVM estimators, we first need to determine the groups. To do this, we used the GICS sector (consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, telecommunications services, and utilities). Figure 2 shows the number of companies in each GICS sector. The smallest group (telecommunications services) and the largest group (financials) have 3 and 36 companies, respectively.

In practice, we need to choose appropriate thresholding levels to regularize the realized volatility matrix estimators. We used the following scheme (Wang and Zou (2010)). It is reasonable to assume that financial time series at daily level are stationary during a reasonable time period, and so we employed one-day-ahead prediction error as a performance measure for the prediction evaluation. Denote by  $\hat{\Gamma}_d$ ,  $d = 1, \dots, 63$ , the

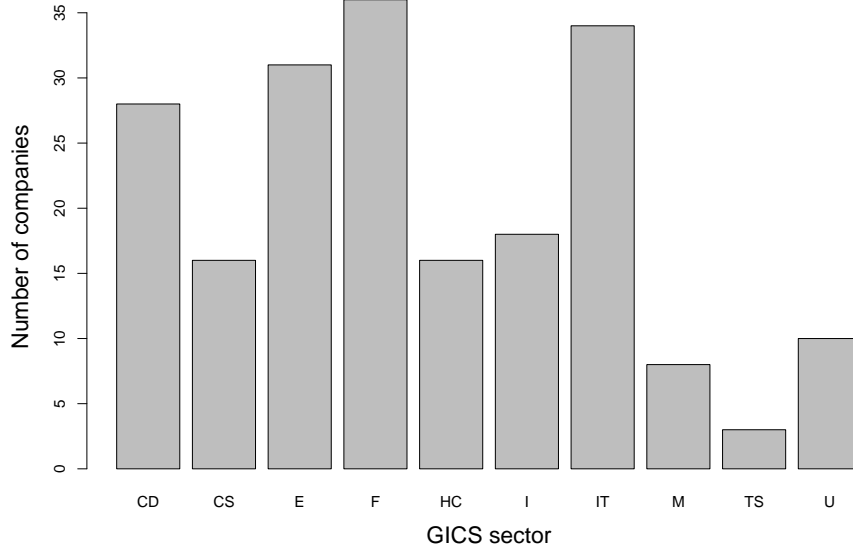
**Table 1.** The MEs of MSRV, KRV, PRV, and PHYV estimators, and their corresponding thresholded estimators and BRV estimators for  $p = 256$  and  $n^{all} = 2000$ .

Type	ME	Noise level	None	Regularizing		BRV		BRV II	
			None	Hard	Soft	Hard	Soft	Hard	Soft
MSRV	Weighted	0.001	2.612	2.612	2.604	1.366	1.285	1.778	1.684
		0.01	2.680	2.680	2.671	1.401	1.310	1.819	1.714
		0.02	2.890	2.890	2.880	1.457	1.368	1.926	1.815
		0.05	3.867	3.867	3.845	1.907	1.770	2.528	2.374
		0.1	5.244	5.244	5.180	2.825	2.644	3.602	3.414
	Spectral	0.001	51.786	51.786	43.833	41.161	41.239	44.209	44.311
		0.01	52.209	52.209	43.707	41.271	41.323	44.340	44.412
		0.02	52.488	52.487	44.636	41.637	41.658	44.731	44.768
		0.05	54.136	54.136	46.062	42.033	42.028	45.354	45.346
		0.1	60.395	60.395	51.907	47.301	47.270	50.848	50.788
PARV	Weighted	0.001	1.958	1.958	1.958	1.849	1.673	2.156	1.914
		0.01	1.962	1.962	1.962	1.812	1.642	2.150	1.900
		0.02	2.061	2.061	2.061	1.753	1.604	2.178	1.913
		0.05	3.070	3.070	3.069	1.469	1.326	2.017	1.861
		0.1	6.981	6.981	6.970	2.262	1.566	2.728	2.139
	Spectral	0.001	25.854	25.854	24.764	21.727	21.752	24.188	24.215
		0.01	25.835	25.835	24.732	21.486	21.535	24.015	24.056
		0.02	26.243	26.243	25.109	21.260	21.438	24.039	24.088
		0.05	27.795	27.795	26.798	20.404	20.851	22.892	23.356
		0.1	33.307	33.307	32.410	22.203	22.346	24.857	25.002
KRV	Weighted	0.001	1.822	1.822	1.822	1.770	0.897	1.734	1.162
		0.01	1.874	1.874	1.873	1.780	0.912	1.757	1.184
		0.02	2.061	2.061	2.060	1.811	0.964	1.838	1.260
		0.05	3.587	3.587	3.585	1.459	1.287	1.941	1.819
		0.1	6.692	6.692	6.641	2.822	2.669	3.833	3.641
	Spectral	0.001	32.803	32.803	30.183	25.251	25.540	27.595	27.867
		0.01	32.663	32.663	29.869	24.948	25.221	27.310	27.553
		0.02	33.546	33.546	30.901	25.726	25.999	28.202	28.403
		0.05	35.235	35.235	32.330	26.049	26.269	28.746	28.982
		0.1	42.112	42.111	39.587	31.026	30.952	34.242	34.127
PHYV	Weighted	0.001	2.589	2.589	2.585	1.099	1.099	1.435	1.477
		0.01	2.594	2.594	2.590	1.101	1.100	1.436	1.474
		0.02	2.618	2.618	2.614	1.098	1.104	1.430	1.488
		0.05	2.813	2.813	2.809	1.096	1.119	1.452	1.517
		0.1	4.159	4.159	4.154	1.124	1.205	1.628	1.759
	Spectral	0.001	36.748	36.748	32.629	28.474	28.598	30.904	31.027
		0.01	37.520	37.520	32.913	29.151	29.239	31.617	31.701
		0.02	37.565	37.565	33.439	29.416	29.546	31.875	31.996
		0.05	37.509	37.509	32.832	28.965	29.031	31.446	31.491
		0.1	40.296	40.296	35.674	30.780	30.905	33.361	33.560

realized volatility matrix estimators being one of MSRV, PARV, and KRV. The mean squared prediction error (MSPE) is defined as

$$MSPE_R(\mathcal{T}_\lambda[\hat{\Gamma}]) = \frac{1}{62} \sum_{d=1}^{62} \|\mathcal{T}_\lambda[\hat{\Gamma}_d] - \hat{\Gamma}_{d+1}\|_2^2 \quad \text{and}$$

$$MSPE_{BR}(\mathcal{T}_\lambda[\hat{\Sigma}]) = \frac{1}{62} \sum_{d=1}^{62} \|\hat{\Theta}_d + \mathcal{T}_\lambda[\hat{\Sigma}_d] - \hat{\Gamma}_{d+1}\|_2^2,$$



**Figure 2.** The number of companies in each GICS sector: consumer discretionary (CD), consumer staples (CS), energy (E), financials (F), health care (HC), industrials (I), information technology (IT), materials (M), telecommunications services (TS), utilities (U).

where  $\lambda$  is the tuning parameter,  $\hat{\Theta}_d, d = 1, \dots, 63$ , are factor volatility matrix estimators,  $\hat{\Sigma}_d = \hat{\Gamma}_d - \hat{\Theta}_d, d = 1, \dots, 63$ , are sparse non-factor volatility matrix estimators, and the thresholding function  $\mathcal{T}_\lambda$  can be the adaptive thresholding functions,  $\mathcal{T}_\omega^H$  and  $\mathcal{T}_\omega^S$  in (3.4). We minimized MSPE to select the thresholding levels, and the tuning parameters  $M = N = \lfloor n^{1/2} \rfloor, K = \lfloor n^{1/2} \rfloor, J = \lfloor n^{1/4} \rfloor$  and  $H = \lfloor n^{1/2} \rfloor$  for MSRVM, PARVM, and KRVM, respectively.

**Table 2.** The MSPEs of MSRVM, KRVM, PARVM, KRVP and PARVPM estimators, and their corresponding thresholded estimators. The relative efficiencies with respect to realized volatility estimators without thresholding are given in the parentheses.

	MSPE $\times 10^2$					
	None	BRVM			Regularizing	
	None	Hard	Soft	Hard	Soft	
MSRVM	0.6552 (1.000)	0.5998 (1.0923)	0.5987 (1.0943)	0.6384 (1.0263)	0.6137 (1.0676)	
PARVM	0.5214 (1.000)	0.4810 (1.0839)	0.4803 (1.0856)	0.5183 (1.0060)	0.4977 (1.0476)	
KRVM	0.6878 (1.000)	0.6262 (1.0984)	0.6270 (1.0971)	0.6661 (1.0327)	0.6397 (1.0752)	

Table 2 reports the MSPE results for BRVM estimators and regularizing estimators based on MSRVM, PARVM, and KRVM. We found that the BRVM estimators have the best performance. Specifically, the BRVM estimators have 8%-10% less MSPE than their

corresponding realized volatility matrix estimators without regularizing nor blocking. On the other hand, comparing the realized volatility matrix estimators, and the corresponding adaptive thresholding estimators, we found that regularizing only can make some reduction on MSPE. Regarding the realized volatility matrix estimation methods (MSRVM, PARVM, and KRVM), the PARVM estimators have the smallest MSPE. These findings may indicate some heterogeneity and block structures in the volatility matrices of the 200 chosen stocks.

## 7. Concluding remarks

This paper investigates large volatility matrix estimation for high-frequency financial data. We consider factors in high-frequency asset pricing model and incorporate the factor influences into the proposed large volatility matrix estimation by the blocking and regularizing approach. The main idea is that assets' prices are governed by some common factors, and that assets with similar characteristics share the same association with the factors. Under the proposed factor-based model, we develop an estimation scheme called "blocking and regularizing". The proposed BRVM estimator shows advantage over thresholded MSRVM, PARVM and KRVM estimators in the presence of factor influence, both theoretically and numerically. One issue in this study is that correctly identifying the blocking structure of the volatility matrix plays an important role in large volatility matrix estimation, and more research is needed in future to revolve this issue. A future research direction is to build some links between models and data at high-frequency and low-frequency levels and develop the joint inference for the combined data (see Kim and Wang (2016)).

## 8. Proofs

Let  $\mathbf{x} = (x_1, \dots, x_p)^T$  be a  $p$ -dimensional vector and  $\mathbf{U} = (U_{ij})$  be a  $p$  by  $p$  matrix.  $\ell_d$ -norms for  $\mathbf{x}$  and  $\mathbf{U}$  are defined as follows, respectively,

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

Under the definitions, matrix spectral norm  $\|\mathbf{U}\|_2$  is equivalent to the square root of largest eigenvalue of  $\mathbf{U}\mathbf{U}^T$ , and

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

For a symmetric matrix  $\mathbf{U}$ , its matrix norm is equal to the largest absolute eigenvalue, and  $\|\mathbf{U}\|_2 \leq \|\mathbf{U}\|_1 = \|\mathbf{U}\|_\infty$ . In this paper, we consider  $\ell_d$  norms for  $d = 1, 2, \infty$ .

To ease the notations, we assume that  $n$  is given.



### 8.1. Proof of Theorem 4.1

**Proof of Theorem 4.1.** First we derive the element-wise bound for  $\Theta$  estimator and  $\Sigma$  estimator constructed from a realized volatility matrix  $\widehat{\Gamma}$ . By the construction of  $\Theta$  in (2.3), we have

$$\begin{aligned}
& \mathbb{E} \left[ |\widehat{\theta}_{ab} - \theta_{ab}|^\beta \right] \\
&= \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} (\widehat{\Gamma}_{ij} - \theta_{ab}) \right|^\beta \right] \\
&= \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} (\widehat{\Gamma}_{ij} - \Gamma_{ij}) + \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} (\Gamma_{ij} - \theta_{ab}) \right|^\beta \right] \\
&\leq C \left\{ \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} (\widehat{\Gamma}_{ij} - \Gamma_{ij}) \right|^\beta \right] + \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} (\Gamma_{ij} - \theta_{ab}) \right|^\beta \right] \right\} \\
&\leq C \left\{ \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} \mathbb{E} \left[ |\widehat{\Gamma}_{ij} - \Gamma_{ij}|^\beta \right] + \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} \Sigma_{ij} \right|^\beta \right] \right\}. \tag{8.1}
\end{aligned}$$

For the second term in (8.1), we have

$$\begin{aligned}
& \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} \Sigma_{ij} \right|^\beta \right] \\
&\leq \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} |\Sigma_{ij}|^\delta |\Sigma_{ij}|^{1-\delta} \right|^\beta \right] \\
&\leq \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \sum_{j=b_1}^{b_{p_b}} |\Sigma_{ij}|^\delta |\Sigma_{ii} \Sigma_{jj}|^{(1-\delta)/2} \right|^\beta \right] \\
&\leq \mathbb{E} \left[ \left| \frac{1}{p_a p_b} \sum_{i=a_1}^{a_{p_a}} \Psi \frac{\pi(p)}{B} \right|^\beta \right] \\
&\leq \mathbb{E} \left[ \left| \Psi \frac{\pi(p)}{B p_b} \right|^\beta \right] \leq C \mathbb{E} \left[ \Psi^\beta (\pi(p)/p)^\beta \right] \leq C (\pi(p)/p)^\beta, \tag{8.2}
\end{aligned}$$

where the third inequality is due to (3.2).

Collecting (4.2), (8.1), and (8.2), we have

$$\mathbb{E} \left[ |\widehat{\theta}_{ab} - \theta_{ab}|^\beta \right] \leq C \left( n^{-\beta/4} + (\pi(p)/p)^\beta \right). \tag{8.3}$$

For the elements in  $\widehat{\Sigma}$ , we have

$$\begin{aligned}\mathbb{E} \left[ |\widehat{\Sigma}_{ij} - \Sigma_{ij}|^\beta \right] &\leq \mathbb{E} \left[ |\widehat{\Gamma}_{ij} - \Gamma_{ij}|^\beta \right] + \mathbb{E} \left[ |\widehat{\Theta}_{ij} - \Theta_{ij}|^\beta \right] \\ &\leq C \left( n^{-\beta/4} + (\pi(p)/p)^\beta \right),\end{aligned}$$

where the last inequality is due to (4.2) and (8.3). ■

## 8.2. Proof of Theorem 4.2

**Proposition 1.** *Under the assumptions in Theorem 4.2, we have for  $\widetilde{\Sigma}$*

$$\mathbb{E} \left[ \|\widetilde{\Sigma} - \Sigma\|_2 \mid n \right] \leq \mathbb{E} \left[ \|\widetilde{\Sigma} - \Sigma\|_\infty \mid n \right] \leq C\pi(p)(p^{2/\beta}e_n)^{1-\delta} \text{ a.s.}$$

**Proof of Proposition 1.** Because of similarity, we provide arguments only for the hard threshold rule. Since  $|\widehat{\Sigma}_{ii} \vee 0 - \Sigma_{ii}| \leq |\widehat{\Sigma}_{ii} - \Sigma_{ii}|$ , without loss of generality we assume that  $\widehat{\Sigma}_{ii} \geq 0$ . Let

$$\begin{aligned}D &= (D_{ij})_{i,j=1,\dots,p}, \quad D_{ij} = (\widetilde{\Sigma}_{ij} - \Sigma_{ij})\mathbf{1}(A_{ij}^c), \quad \text{and} \\ A_{ij} &= \{|\widetilde{\Sigma}_{ij} - \Sigma_{ij}| \leq 2 \min\{|\Sigma_{ij}|, \varpi_{ij}\}\}.\end{aligned}$$

By triangle inequality, we have

$$\mathbb{E} \left[ \|\widetilde{\Sigma} - \Sigma\|_\infty \right] \leq \mathbb{E} \left[ \|\widetilde{\Sigma} - \Sigma - D\|_\infty \right] + \mathbb{E} \left[ \|D\|_\infty \right].$$

First, consider  $\mathbb{E} \left[ \|\widetilde{\Sigma} - \Sigma - D\|_\infty \right]$ . We have

$$\begin{aligned}&\mathbb{E} \left[ \|\widetilde{\Sigma} - \Sigma - D\|_\infty \right] \\ &= \mathbb{E} \left[ \max_{1 \leq i \leq p} \sum_{j=1}^p |\widetilde{\Sigma}_{ij} - \Sigma_{ij}| \mathbf{1}(A_{ij}) \right] \\ &\leq \mathbb{E} \left[ \max_{1 \leq i \leq p} \sum_{j=1}^p 2\varpi_{ij} \mathbf{1}(|\Sigma_{ij}| \geq \varpi_{ij}) \right] \\ &\quad + \mathbb{E} \left[ \max_{1 \leq i \leq p} \sum_{j=1}^p 2|\Sigma_{ij}| \mathbf{1}(|\Sigma_{ij}| < \varpi_{ij}) \right] \\ &\leq 4\mathbb{E} \left[ \max_{1 \leq i \leq p} \sum_{j=1}^p |\Sigma_{ij}|^\delta \varpi_{ij}^{1-\delta} \right] \\ &\leq C\mathbb{E} \left[ \max_{1 \leq i \leq p} \sum_{j=1}^p |\Sigma_{ij}|^\delta (\Sigma_{ii}\Sigma_{jj})^{(1-\delta)/2} (p^{2/\beta}e_n)^{1-\delta} \left[ \sqrt{\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}} / \sqrt{\Sigma_{ii}\Sigma_{jj}} \right]^{1-\delta} \right]\end{aligned}$$

$$\begin{aligned}
&\leq C\mathbb{E}[\Psi^2]^{1/2}\pi(p)(p^{2/\beta}e_n)^{1-\delta}\mathbb{E}\left[\max_{i,j}\left(\frac{\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}}{\Sigma_{ii}\Sigma_{jj}}\right)^{1-\delta}\right]^{1/2} \\
&\leq C\pi(p)(p^{2/\beta}e_n)^{1-\delta},
\end{aligned} \tag{8.4}$$

where the last inequality is due to (8.5) below. We have

$$\begin{aligned}
\mathbb{E}\left[\max_{i,j}\left(\frac{\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}}{\Sigma_{ii}\Sigma_{jj}}\right)^{1-\delta}\right] &= \mathbb{E}\left[\max_{i,j}\left(\frac{\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}}{\Sigma_{ii}\Sigma_{jj}}+1\right)^{1-\delta}\right] \\
&\leq C\mathbb{E}\left[\max_{i,j}\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|^{1-\delta}\right]+C \\
&\leq C,
\end{aligned} \tag{8.5}$$

where the last inequality is due to the fact that we have

$$\begin{aligned}
&\mathbb{E}\left[\max_{i,j}\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\right] \\
&\leq \sum_{i,j=1}^p\mathbb{E}\left[\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\mathbf{1}\left(\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\geq p^{4/\beta}e_n\right)\right] \\
&\quad +\mathbb{E}\left[\max_{i,j}\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\mathbf{1}\left(\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|<p^{4/\beta}e_n\right)\right] \\
&\leq \sum_{i,j=1}^p\mathbb{E}\left[\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\mathbf{1}\left(\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\geq p^{4/\beta}e_n\right)\right]+p^{4/\beta}e_n \\
&\leq \sum_{i,j=1}^p\mathbb{E}\left[\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|^{\beta/2}\right]^{2/\beta}P\left(\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\geq p^{4/\beta}e_n\right)^{1-2/\beta}+p^{4/\beta}e_n \\
&\leq Cp^{4/\beta}e_n,
\end{aligned} \tag{8.6}$$

where the third inequality is due to Hölder's inequality, and the last inequality is due to (8.7) and (8.8) below. We have for  $x=p^{4/\beta}e_n$ ,

$$\begin{aligned}
P\left(\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|\geq x\right) &\leq \mathbb{E}\left[\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|^{\beta/2}\right]/x^{\beta/2} \\
&\leq Ce_n^{\beta/2}/x^{\beta/2}\leq Cp^{-2},
\end{aligned} \tag{8.7}$$

where the first inequality is due to Chebyshev inequality, and the second inequality holds by the fact that

$$\begin{aligned}
&\mathbb{E}\left[\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}-\Sigma_{ii}\Sigma_{jj}\right|^{\beta/2}\right] \\
&\leq \mathbb{E}\left[\left|(\widehat{\Sigma}_{ii}-\Sigma_{ii})\widehat{\Sigma}_{jj}+(\widehat{\Sigma}_{jj}-\Sigma_{jj})\Sigma_{ii}\right|^{\beta/2}\right] \\
&\leq C\mathbb{E}\left[\left|(\widehat{\Sigma}_{ii}-\Sigma_{ii})\widehat{\Sigma}_{jj}\right|^{\beta/2}\right]+C\mathbb{E}\left[\left|(\widehat{\Sigma}_{jj}-\Sigma_{jj})\Sigma_{ii}\right|^{\beta/2}\right]
\end{aligned}$$

$$\begin{aligned}
&\leq C\sqrt{\mathbb{E}\left[|\widehat{\Sigma}_{ii} - \Sigma_{ii}|^\beta\right] \mathbb{E}\left[|\widehat{\Sigma}_{jj}|^\beta\right]} + C\sqrt{\mathbb{E}\left[|\widehat{\Sigma}_{jj} - \Sigma_{jj}|^\beta\right] \mathbb{E}\left[|\Sigma_{ii}|^\beta\right]} \\
&\leq Ce_n^{\beta/2}, \tag{8.8}
\end{aligned}$$

where the third and fourth inequalities are due to Hölder's inequality and Theorem 4.1, respectively.

Consider  $\mathbb{E}[\|\mathbf{D}\|_\infty]$ . Simple algebra shows

$$\begin{aligned}
\mathbb{E}[\|\mathbf{D}\|_\infty] &\leq \mathbb{E}\left[\max_{1 \leq i \leq p} \sum_{j=1}^p |\Sigma_{ij}| \mathbf{1}(A_{ij}^c) \mathbf{1}(|\widehat{\Sigma}_{ij}| < \varpi_{ij})\right] \\
&\quad + \mathbb{E}\left[\max_{1 \leq i \leq p} \sum_{j=1}^p |\widehat{\Sigma}_{ij} - \Sigma_{ij}| \mathbf{1}(A_{ij}^c) \mathbf{1}(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij})\right] \\
&= (I) + (II).
\end{aligned}$$

For (I), we have

$$\begin{aligned}
(I) &\leq \mathbb{E}\left[\max_{1 \leq i \leq p} \sum_{j=1}^p |\Sigma_{ij} - \widehat{\Sigma}_{ij}| \mathbf{1}(A_{ij}^c) \mathbf{1}(|\widehat{\Sigma}_{ij}| < \varpi_{ij})\right] \\
&\quad + \mathbb{E}\left[\max_{1 \leq i \leq p} \sum_{j=1}^p \varpi_{ij} \mathbf{1}(A_{ij}^c) \mathbf{1}(|\widehat{\Sigma}_{ij}| < \varpi_{ij})\right] \\
&\leq \sum_{i,j=1}^p \mathbb{E}\left[|\Sigma_{ij} - \widehat{\Sigma}_{ij}| \mathbf{1}(A_{ij}^c) \mathbf{1}(|\widehat{\Sigma}_{ij}| < \varpi_{ij})\right] \\
&\quad + \mathbb{E}\left[\max_{1 \leq i \leq p} \sum_{j=1}^p \varpi_{ij} \mathbf{1}(A_{ij}^c) \mathbf{1}(|\widehat{\Sigma}_{ij}| < \varpi_{ij})\right] \\
&\leq \sum_{i,j=1}^p \mathbb{E}\left[|\Sigma_{ij} - \widehat{\Sigma}_{ij}|^\beta\right]^{1/\beta} P\left(|\widehat{\Sigma}_{ij} - \Sigma_{ij}| > \varpi_{ij}\right)^{1-1/\beta} \\
&\quad + \mathbb{E}\left[\max_{1 \leq i \leq p} \sum_{j=1}^p \varpi_{ij} \mathbf{1}(|\Gamma_{ij}| > 2\varpi_{ij})\right] \\
&\leq Ce_n p^{2/\beta} + \mathbb{E}\left[\max_{1 \leq i \leq p} \sum_{j=1}^p \varpi_{ij} \mathbf{1}(|\Gamma_{ij}| > 2\varpi_{ij})\right] \\
&\leq Ce_n p^{2/\beta} + C\pi(p)(p^{2/\beta} e_n)^{1-\delta}, \tag{8.9}
\end{aligned}$$

where the fourth inequality is due to Theorem 4.1 and (8.10) below, and the last inequality is derived similar to the proof of (8.4). We have for large  $n$ ,

$$P\left(|\widehat{\Sigma}_{ij} - \Sigma_{ij}| > \varpi_{ij}\right)$$

$$\begin{aligned}
&\leq P\left(|\widehat{\Sigma}_{ij} - \Sigma_{ij}| > \varpi_{ij}, \left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj} - \Sigma_{ii}\Sigma_{jj}\right| \leq p^{4/\beta}e_n\right) \\
&\quad + P\left(\left|\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj} - \Sigma_{ii}\Sigma_{jj}\right| \geq p^{4/\beta}e_n\right) \\
&\leq P\left(|\widehat{\Sigma}_{ij} - \Sigma_{ij}| > \frac{1}{2}C_{\Sigma}\lambda\right) + Cp^{-2} \\
&\leq Cp^{-2}, \tag{8.10}
\end{aligned}$$

where the second inequality is due to (8.7), and the last inequality can be derived similar to the proof of (8.7).

For (II),

$$\begin{aligned}
(II) &\leq \sum_{i,j=1}^p \mathbb{E}\left[|\widehat{\Sigma}_{ij} - \Sigma_{ij}|\mathbf{1}(A_{ij}^c)\mathbf{1}(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij})\right] \\
&\leq \sum_{i,j=1}^p \mathbb{E}\left[|\widehat{\Sigma}_{ij} - \Sigma_{ij}|^{\beta}\right]^{1/\beta} P\left(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}, A_{ij}^c\right)^{1-1/\beta} \\
&\leq \sum_{i,j=1}^p e_n P\left(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}, A_{ij}^c\right)^{1-1/\beta} \\
&\leq Ce_n p^{2/\beta}, \tag{8.11}
\end{aligned}$$

where the third inequality is due to Theorem 4.1, and the last inequality holds by the fact that

$$\begin{aligned}
&P\left(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}, A_{ij}^c\right) \\
&= P\left(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}, A_{ij}^c, |\Sigma_{ij}| \geq \varpi_{ij}\right) + P\left(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}, A_{ij}^c, \varpi_{ij}/2 \leq |\Sigma_{ij}| < \varpi_{ij}\right) \\
&\quad + P\left(|\widehat{\Sigma}_{ij}| \geq \varpi_{ij}, A_{ij}^c, |\Sigma_{ij}| < \varpi_{ij}/2\right) \\
&\leq P\left(|\widehat{\Sigma}_{ij} - \Sigma_{ij}| > 2\varpi_{ij}\right) + P\left(|\widehat{\Sigma}_{ij} - \Sigma_{ij}| > \varpi_{ij}\right) + P\left(|\widehat{\Sigma}_{ij} - \Sigma_{ij}| > \varpi_{ij}/2\right) \\
&\leq Cp^{-2},
\end{aligned}$$

where the last inequality can be derived similar to the proof of (8.10).

The statement follows from (8.4), (8.9), and (8.11). ■

**Proof of Theorem 4.2.** We have

$$\begin{aligned}
\|\widetilde{\Gamma} - \Gamma\|_{\Gamma} &\leq \|\widehat{\Theta} - \Theta\|_{\Gamma} + \|\widetilde{\Sigma} - \Sigma\|_{\Gamma} \\
&= I + II.
\end{aligned}$$

First, consider  $I$ . We have the following eigen-decompositions,

$$\Theta = U\Xi U^T \quad \text{and} \quad \widehat{\Theta} = \widehat{U}\widehat{\Xi}\widehat{U}^T,$$

where  $\Xi = \text{diag}(\lambda_1(\Theta), \dots, \lambda_B(\Theta))$ ,  $\widehat{\Xi} = \text{diag}(\lambda_1(\widehat{\Theta}), \dots, \lambda_B(\widehat{\Theta}))$ , and  $U, \widehat{U} \in \mathbb{R}^{p \times B}$  are formed by eigenvectors of  $\Theta$  and  $\widehat{\Theta}$ , respectively. Define

$$H = U\Xi^{1/2} \quad \text{and} \quad \widehat{H} = \widehat{U}\widehat{\Xi}^{1/2}.$$

Then we have

$$\begin{aligned}\widehat{\mathbf{H}} &= \mathbf{U}\boldsymbol{\Xi}^{1/2}\boldsymbol{\Xi}^{-1/2}\widehat{\boldsymbol{\Xi}}^{1/2} + (\widehat{\mathbf{U}} - \mathbf{U})\widehat{\boldsymbol{\Xi}}^{1/2} \\ &= \mathbf{H}\boldsymbol{\Xi}^{-1/2}\widehat{\boldsymbol{\Xi}}^{1/2} + (\widehat{\mathbf{U}} - \mathbf{U})\widehat{\boldsymbol{\Xi}}^{1/2}.\end{aligned}$$

Simple algebraic manipulations show

$$\begin{aligned}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{\Gamma} &\leq \|\mathbf{H}(\boldsymbol{\Xi}^{-1/2}\widehat{\boldsymbol{\Xi}}\boldsymbol{\Xi}^{-1/2} - \mathbf{I}_B)\mathbf{H}^T\|_{\Gamma} + \|(\widehat{\mathbf{U}} - \mathbf{U})\widehat{\boldsymbol{\Xi}}(\widehat{\mathbf{U}} - \mathbf{U})^T\|_{\Gamma} \\ &\quad + 2\|\mathbf{H}\boldsymbol{\Xi}^{-1/2}\widehat{\boldsymbol{\Xi}}(\widehat{\mathbf{U}} - \mathbf{U})^T\|_{\Gamma} \\ &= G_1 + G_2 + 2G_3.\end{aligned}$$

For  $G_1$ , we have

$$\begin{aligned}G_1^2 &= p^{-1}\|\boldsymbol{\Gamma}^{-1/2}\mathbf{H}(\boldsymbol{\Xi}^{-1/2}\widehat{\boldsymbol{\Xi}}\boldsymbol{\Xi}^{-1/2} - \mathbf{I}_B)\mathbf{H}^T\boldsymbol{\Gamma}^{-1/2}\|_F^2 \\ &\leq p^{-1}\|\boldsymbol{\Xi}^{-1/2}\widehat{\boldsymbol{\Xi}}\boldsymbol{\Xi}^{-1/2} - \mathbf{I}_B\|_F^2\|\mathbf{H}^T\boldsymbol{\Gamma}^{-1}\mathbf{H}\|_2^2 \\ &= p^{-1}\sum_{j=1}^B\left(\frac{\lambda_j(\widehat{\boldsymbol{\Theta}}) - \lambda_j(\boldsymbol{\Theta})}{\lambda_j(\boldsymbol{\Theta})}\right)^2\|\mathbf{H}^T\boldsymbol{\Gamma}^{-1}\mathbf{H}\|_2^2 \\ &\leq p^{-1}\lambda_B(\boldsymbol{\Theta})^{-2}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_F^2\|\mathbf{H}^T\boldsymbol{\Gamma}^{-1}\mathbf{H}\|_2^2 \\ &\leq 4p^{-1}\lambda_B(\boldsymbol{\Theta})^{-2}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_F^2 \\ &\leq 4p^{-1}\frac{Bc_2^2}{\|\boldsymbol{\Theta}\|_F^2}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_F^2 \\ &\leq \frac{CB}{p^3}\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_F^2 \text{ a.s.},\end{aligned}\tag{8.12}$$

where the second and third inequalities are due to Weyl's Theorem (Li (1998a)) and (8.13), respectively, and the fourth and last inequalities are due to Assumption 5. By the Sherman-Morrison-Woodbury formula, we obtain

$$\boldsymbol{\Gamma}^{-1} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{H}[\mathbf{I}_B + \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}]^{-1}\mathbf{H}^T\boldsymbol{\Sigma}^{-1}.$$

Then we have

$$\begin{aligned}\mathbf{H}^T\boldsymbol{\Gamma}^{-1}\mathbf{H} &= \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H} - \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}[\mathbf{I}_B + \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}]^{-1}\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H} \\ &= \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}[\mathbf{I}_B + \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}]^{-1} \\ &= \mathbf{I}_B - [\mathbf{I}_B + \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}]^{-1}.\end{aligned}$$

Since  $\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}$  is positive semi-definite, we have

$$\|\mathbf{H}^T\boldsymbol{\Gamma}^{-1}\mathbf{H}\|_2 \leq 1 + \|[\mathbf{I}_B + \mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}]^{-1}\|_2 \leq 2.\tag{8.13}$$

For  $G_2$ , since we have

$$\inf_{\mathcal{O} \in \mathcal{G}_{B \times B}} \|\widehat{\mathbf{U}} - \mathbf{U}\mathcal{O}\|_F^2 \leq 2\|\sin(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Theta})\|_F^2,$$

where  $\mathcal{G}_{B \times B} = \{\mathcal{O} \in \mathcal{V}_{B \times B}, \mathbf{U}\mathcal{O}\Xi\mathcal{O}^T\mathbf{U}^T = \Theta\}$ ,  $\mathcal{V}_{B \times B} = \{\mathcal{O} \in \mathbb{R}^{B \times B}, \mathcal{O}\mathcal{O}^T = \mathbf{I}_B\}$ , and  $\|\sin(\widehat{\Theta}, \Theta)\|_F^2 = \|\widehat{\Theta}\widehat{\Theta}^T(\mathbf{I}_p - \Theta\Theta^T)\|_F^2$ , without loss of generality we assume

$$\|\widehat{\mathbf{U}} - \mathbf{U}\|_F^2 = \inf_{\mathcal{O} \in \mathcal{G}_{B \times B}} \|\widehat{\mathbf{U}} - \mathbf{U}\mathcal{O}\|_F^2.$$

Then, by Davis-Kahan's  $\sin \theta$  theorem (Li (1998b)), we have

$$\|\widehat{\mathbf{U}} - \mathbf{U}\|_F^2 \leq 2 \frac{\|(\widehat{\Theta} - \Theta)\mathbf{U}\|_F^2}{\lambda_B(\Theta)^2}. \quad (8.14)$$

Thus, we have

$$\begin{aligned} G_2 &\leq p^{-1/2} \|\widehat{\mathbf{U}} - \mathbf{U}\|_F^2 \|\widehat{\Xi}\|_2 \|\Gamma^{-1}\|_2 \\ &\leq Cp^{-1/2} \lambda_B(\Theta)^{-2} \|\widehat{\Theta} - \Theta\|_F^2 \|\widehat{\Xi}\|_2 \\ &\leq C \frac{B}{p^{5/2}} \|\widehat{\Theta} - \Theta\|_F^2 \|\widehat{\Xi}\|_2. \end{aligned} \quad (8.15)$$

For  $G_3$ , we have

$$\begin{aligned} G_3^2 &= \|\mathbf{H}\Xi^{-1/2}\widehat{\Xi}(\widehat{\mathbf{U}} - \mathbf{U})^T\|_{\Gamma}^2 \\ &\leq p^{-1} \|\widehat{\mathbf{U}} - \mathbf{U}\|_F^2 \|\Xi^{-1/2}\widehat{\Xi}\|_2^2 \|\mathbf{H}^T\Gamma^{-1}\mathbf{H}\|_2 \|\Gamma^{-1}\|_2 \\ &\leq Cp^{-1} \lambda_B(\Theta)^{-3} \|\widehat{\Theta} - \Theta\|_F^2 \|\widehat{\Xi}\|_2^2 \\ &\leq C \frac{B^{3/2}}{p^4} \|\widehat{\Theta} - \Theta\|_F^2 \|\widehat{\Xi}\|_2^2, \end{aligned} \quad (8.16)$$

where the second inequality is due to (8.14).

We have

$$\begin{aligned} E \left[ \|\widehat{\Xi}\|_2 \right] &\leq E \left[ \|\Xi\|_2 \right] + E \left[ \|\widehat{\Xi} - \Xi\|_2 \right] \\ &\leq E \left[ \|\widehat{\Theta} - \Theta\|_2 \right] + Cp/B^{1/2} \\ &\leq E \left[ \max_{1 \leq i \leq p} \sum_{j=1}^p |\widehat{\Theta}_{ij} - \Theta_{ij}| \right] + Cp/B^{1/2} \\ &\leq B \max_{1 \leq i \leq p} \sum_{j=1}^p E \left[ |\widehat{\Theta}_{ij} - \Theta_{ij}| \right] + Cp/B^{1/2} \\ &\leq Bpe_n + Cp/B^{1/2}, \end{aligned} \quad (8.17)$$

where the last inequality is due to Theorem 4.1. Collecting (8.12), (8.15), (8.16), and (8.17), we obtain

$$\begin{aligned} \|\widehat{\Theta} - \Theta\|_{\Gamma} &\leq C \left( \frac{B^{1/2}}{p^{3/2}} \|\widehat{\Theta} - \Theta\|_F + \frac{B}{p^{5/2}} \|\widehat{\Theta} - \Theta\|_F^2 \|\widehat{\Xi}\|_2 + \frac{B^{3/4}}{p^2} \|\widehat{\Theta} - \Theta\|_F \|\widehat{\Xi}\|_2 \right) \\ &\leq O_p \left( [B^{1/2}p^{1/2}e_n^2 + B^{1/4}e_n][1 + B^{3/2}e_n] \right). \end{aligned} \quad (8.18)$$

Consider *II*. By Proposition 1, we have

$$\begin{aligned} \|\tilde{\Sigma} - \Sigma\|_{\Gamma}^2 &= p^{-1} \|(\tilde{\Sigma} - \Sigma)\Gamma^{-1}\|_F^2 \\ &\leq p^{-1} \|\tilde{\Sigma} - \Sigma\|_2^2 \|\Gamma^{-1}\|_F^2 \\ &\leq O_p(\pi(p)^2 (p^{2/\beta} e_n)^{2-2\delta}). \end{aligned} \tag{8.19}$$

The statement is proved by (8.18) and (8.19). ■

**Acknowledgments:** The research of Yazhen Wang was supported in part by NSF grants DMS-105635, DMS-1265203 and DMS-1528375.

## References

- [1] Aguilar, O., and West, M. (2000), “Bayesian Dynamic Factor Models and Portfolio Allocation,” *Journal of Business & Economic Statistics*, 18, 338–357.
- [2] Ait-Sahalia, Y., Fan, J., and Xiu, D. (2010), “High-frequency covariance estimates with noisy and asynchronous financial data,” *Journal of the American Statistical Association*, 105, 1504–1517.
- [3] Ait-Sahalia, Y. and Xiu, D. (2015). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. Chicago Booth Research Paper, (15–43).
- [4] Anderson, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003), “Modeling and forecasting realized volatility,” *Econometrica*, 71, 579–625.
- [5] Barndorff-Nielsen, O. E., and Shephard, N. (2002), “Econometric analysis of realized volatility and its use in estimating stochastic volatility models,” *Journal of the Royal Statistical Society, Ser. B*, 64, 253–280.
- [6] Barndorff-Nielsen, O. E., and Shephard, N. (2006), “Econometrics of testing for jumps in financial econometrics using bipower variation,” *Journal of Financial Econometrics*, 4, 1–30.
- [7] 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, 1481–536.
- [8] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009), “Realized kernels in practice: Trades and quotes,” *The Econometrics Journal*, 12, C1–C32.
- [9] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (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.
- [10] Bibinger, M., Hautsch, N., Malec, P., and Reiß, M. (2014), “Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency,” *The Annals of Statistics*, 42(4), 1312–1346.
- [11] Bickel, P. J., and Levina, E. (2008a), “Regularized estimation of large covariance matrices,” *The Annals of Statistics*, 36, 199–227.
- [12] Bickel, P. J., and Levina, E. (2008b), “Covariance regularization by thresholding,” *The Annals of Statistics*, 36, 2577–2604.



- [13] Cai, T., and Liu, W. (2011), “Adaptive thresholding for sparse covariance matrix estimation,” *Journal of the American Statistical Association*, 106, 672–684.
- [14] Cai, T., and Zhou, H. (2012), “Optimal rates of convergence for sparse covariance matrix estimation,” *The Annals of Statistics*, 40, 2389–2420.
- [15] Chamberlain, G. (1983), “Funds, Factors, and Diversification in Arbitrage Pricing Models,” *Econometrica*, 51, 1305–1323.
- [16] Chamberlain, G., and Rothschild, M. (1983), “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets,” *Econometrica*, 51, 1281–1304.
- [17] 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.
- [18] Christensen, K., Podolskij, M., and Vetter, M. (2013), “On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes,” *Journal of Multivariate Analysis*, 120, 59–84.
- [19] Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), “A theory of the term structure of interest rates,” *Econometrica*, 53, 385–407.
- [20] Diebold, F. X., and Nerlove, M. (1989), “The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model,” *Journal of Applied Econometrics*, 4, 1–21.
- [21] Engle, R. F., and Watson, M. W. (1981), “A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates,” *Journal of American Statistical Association*, 76, 774–781.
- [22] Fama, E. F., and French, K. R. (1992), “The Cross-section of Expected Stock Returns,” *The Journal of Finance*, 47, 427–465.
- [23] Fama, E. F., and French, K. R. (1993), “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics*, 33, 3–56.
- [24] Fan, J., Fan, Y., and Lv, J. (2008), “High dimensional covariance matrix estimation using a factor model,” *Journal of Econometrics*, 147, 186–197.
- [25] Fan, J., Liao, Y., and Mincheva, M. (2013), “Large covariance estimation by thresholding principal orthogonal complements,” *Journal of the Royal Statistical Society, Ser. B*, 75, 603–680.
- [26] Fan, J., Furger, A., and Xiu, D. (2016), “Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data,” *Journal of Business & Economic Statistics*, 34, 489–503.
- [27] Fan, J., and Wang, Y. (2007), “Multi-scale jump and volatility analysis for high-frequency financial data,” *Journal of American Statistical Association*, 102, 1349–1362.
- [28] Hayashi, T., and Yoshida, N. (2005), “On covariance estimation of non-synchronously observed diffusion processes,” *Bernoulli*, 11, 359–379.
- [29] Huang, X., and Tauchen, G. (2005), “The relative contribution of jumps to total price variance,” *Journal of Financial Econometrics*, 3, 456–499.
- [30] Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009), “Microstructure noise in the continuous case: The Pre-Averaging Approach,” *Stochastic Processes*

- and Their Applications, 119, 2249–2276.
- [31] Kim, D. and Wang, Y. (2016). “Unified discrete-time and continuous-time models and statistical inference for merged low-frequency and high-frequency data,” *Journal of Econometrics* 194, 220-230.
  - [32] Kim, D., Wang, Y., and Zou, J. (2016), “Asymptotic theory for large volatility matrix estimation based on high-frequency financial Data,” *Stochastic Processes and Their Applications*, 126, 3527–3577.
  - [33] Kim, D. Kong, X., Li, C., and Wang, Y. (2015), “Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data,” *Manuscript*.
  - [34] Li, R.-C. (1998a), “Relative perturbation theory: I. Eigenvalue and singular value variations,” *SIAM Journal on Matrix Analysis and Applications*, 19, 956-982.
  - [35] Li, R.-C. (1998b), “Relative perturbation theory: II. Eigenspace and singular subspace variations,” *SIAM Journal on Matrix Analysis and Applications*, 20, 471-492.
  - [36] Mancino, M. E., and Sanfelici, S. (2008), “Robustness of Fourier estimator of integrated volatility in the presence of micro-structure noise,” *Computational Statistics & data analysis*, 52, 2966–2989.
  - [37] Mancino, M. E., and Sanfelici, S. (2011), “Estimating covariance via Fourier method in the presence of asynchronous trading and microstructure noise,” *Journal of Financial Econometrics*, 9(2), 367-408
  - [38] Podolskij, M., and Vetter, M. (2009), “Bipower-type estimation in a noisy diffusion setting,” *Stochastic processes and their applications* ,119, 2803-2831.
  - [39] Ross, S. (1976), “The Theory of Capital Asset Pricing,” *Journal of Economic Theory*, 3, 341–360.
  - [40] Ross, S. (1977), “The Capital Asset Pricing Model (CAMP). Short-Sale Restrictions and Related Issues,” *Journal of Finance*, 32, 177–183.
  - [41] Stock, J. H., and Watson, M. W. (2005), “Implications of dynamic factor models for VAR analysis,” (No. w11467), National Bureau of Economic Research.
  - [42] Tao, M., Wang, Y., and Chen, X. (2013a), “Fast convergence rates in estimating large volatility matrices using high-frequency financial data,” *Econometric Theory*, 29, 838–856.
  - [43] Tao, M., Wang, Y., and Zhou H. (2013b), “Optimal Sparse Volatility Matrix Estimation for High Dimensional Itô Processes With Measurement Errors,” *The Annals of Statistics*, 41(4), 1816–1864.
  - [44] Wang, Y. (2002), “Asymptotic nonequivalence of ARCH models and diffusions,” *The Annals of Statistics*, 30, 754–783.
  - [45] Wang, Y., and Zou, J. (2010), “Vast volatility matrix estimation for high-frequency financial data,” *The Annals of Statistics*, 38, 943–978.
  - [46] Xiu, D. (2010), “Quasi-Maximum Likelihood Estimation of Volatility With High Frequency Data,” *Journal of Econometrics*, 159, 235-250.
  - [47] Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), “A tale of two time scales: Determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association*, 100, 1394–1411.
  - [48] Zhang, L. (2006), “Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach,” *Bernoulli*, 12, 1019–1043.

- [49] Zhang, L. (2011), “Estimating covariation: Epps effect, microstructure noise,” *Journal of Econometrics*, 160, 33-47.