Optimal Covariance Matrix Estimation for High-dimensional Noise in High-frequency Data

Jinyuan Chang[†], Cheng Liu[‡], Cheng Yong Tang[§]

Abstract

In this paper, we consider efficiently learning the structural information from the high-dimensional noise in high-frequency data via estimating its covariance matrix with optimality. The problem is uniquely challenging due to the latency of the targeted high-dimensional vector containing the noises, and the practical reality that the observed data can be highly asynchronous – not all components of the high-dimensional vector are observed at the same time points. To meet the challenges, we propose a new covariance matrix estimator with appropriate localization and thresholding. In the setting with latency and asynchronous observations, we establish the minimax optimal convergence rates associated with two commonly used loss functions for the covariance matrix estimations. As a major theoretical development, we show that despite the latency of the signal in the high-frequency data, the optimal rates remain the same as if the targeted high-dimensional noises are directly observable. Our results indicate that the optimal rates reflect the impact due to the asynchronous observations, which are slower than that with synchronous observations. Furthermore, we demonstrate that the proposed localized estimator with thresholding achieves the minimax optimal convergence rates. We also illustrate the empirical performance of the proposed estimator with extensive simulation studies and a real data analysis.

Keywords: High-dimensional covariance matrix; high-frequency data analysis; measurement error, minimax optimality; thresholding.

[†]School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China. E-mail: changjinyuan@swufe.edu.cn

[‡]School of Economics and Management, Wuhan University, Wuhan, Hubei 430072, China. E-mail: chengliu_eco@whu.edu.cn

[§]Department of Statistical Science, Temple University, Philadelphia, PA 19122-6083, USA. E-mail: yongtang@temple.edu

arXiv:1812.08217v1 [math.ST] 19 Dec 2018

1 Introduction

High-frequency data broadly refer to those collected at time points with very small time intervals between observations. Exemplary scenarios with high-frequency data include the tick-by-tick trading data in finance (Zhang, Mykland and Aït-Sahalia, 2005), longitudinal observations with intensive repeated measurements (Bolger and Laurenceau, 2013), and functional data with dense observations (Zhang and Wang, 2016). High-frequency data are commonly contaminated by some noises, broadly termed as the measurement errors. For measurement errors in the context of functional data analysis, we refer to the review article Wang, Chiou and Müller (2016) and reference therein. In high-frequency financial data, as another example, the microstructure noise is well known; see the monograph Aït-Sahalia and Jacod (2014) for an overview of the problem and methods.

Despite the central interests on recovering the signals contaminated by the noises, the properties of the noises themselves are of their own great interests. For example, Dufrenot, Jawadi and Louhichi (2014) contained studies about the microstructure noise with the context of the financial crises and argued that the microstructure noise can help us understand financial crises; Aït-Sahalia and Yu (2009) investigated the microstructure noise and liquidity measures with high-frequency financial data; Li, Xie and Zheng (2016) found that a parametric function incorporating the market information may account for a substantial contribution to the variations in the microstructure noises. Recently, Jacod, Li and Zheng (2017) highlighted the importance of statistical properties of the microstructure noise and studied the estimation of its moments; Chang et al. (2018) investigated recovering the distribution of the noise with some frequency-domain analysis. Both studies of Jacod, Li and Zheng (2017) and Chang et al. (2018) aimed on univariate case. When the focus turns to multivariate or high-dimensional cases, the covariance between different components of the noise could bring us some useful information for solving practical problems. In the context of aforementioned studies with the noises, additional insights from the covariations between different variables may shed the light on new development of the investigations. For example, they can help on evaluating the magnitude of the noises, identifying the sources contributing to the contamination, and finding some useful patterns in the measurement errors. Nevertheless, studying the covariance between different components of the high-dimensional noise in high-frequency data remains little explored in the literature.

We investigate in this paper the problem of estimating the covariance matrix of the high-dimensional noise in high-frequency data. Our interests are on the validity and optimality of the covariance matrix estimation procedure. The problem is challenging from two major aspects. One aspect is from the fact that the noises of interest are not directly observable, resulting in latency of

the targeted random vectors. We can view the observed data as the contaminated observations of the noises of interest. This challenge arises together with those from the high data dimensionality and high sampling frequency. The properties of high-dimensional covariance matrix estimation have not yet been explored in this important scenario. The other aspect is that the high-dimensional observations may not be synchronous, i.e. different components of the contaminated observation for the high-dimensional noise may be observed at different time points. How this unique data feature affects the statistical properties, more specifically on the validity and optimality, of the covariance matrix estimation also remains unknown.

High-dimensional covariance matrix estimation is an important problem in the current state of knowledge, and has received intensive attentions in the past decade; see, among others, Bickel and Levina (2008a,b), Lam and Fan (2009), Rothman, Levina and Zhu (2009), Cai, Zhang and Zhou (2010), Cai and Liu (2011) and Cai and Zhou (2012a,b). For high-dimensional sparse covariance matrices, the minimax optimality of the estimations were investigated in-depth in Cai and Zhou (2012a,b). We note that the existing estimation methods for high-dimensional sparse covariance matrices are developed when the underlying data of interest are fully observed; hence they are not applicable for the covariance matrix estimation of the noise in high-frequency data with latency and asynchronous observations. In the literature on multivariate and high-dimensional high-frequency data analysis, existing studies mainly concern the estimations of the so-called realized covariance matrix. Specifically, the major objective is on the signal part, attempting to eliminate the impact from the noises; see, for example, Aït-Sahalia, Fan and Xiu (2010), Fan, Li and Yu (2012), Tao, Wang and Zhou (2013), Liu and Tang (2014) and Xia and Zheng (2018). However, it remains little explored on the high-dimensional covariance matrix for the noise in high-frequency data.

Our study makes several important contributions to the area. First, to our knowledge, our method is among the first handling covariance matrix estimation of the high-dimensional noise in high-frequency data. To overcome the difficulty due to the latency and asynchronous observations, we propose a new method with appropriate localization and thresholding. Second, our theoretical analysis establishes the minimax optimal convergence rates associated with two commonly used loss functions for the covariance matrix estimations of the high-dimensional noise in high-frequency data. The minimax optimal rates in this setting are our new theoretical discoveries, and we show that the proposed estimator achieves such rates. Our result also reveals that the optimal convergence rates reflect the impact due to the asynchronous data, which are slower than those with synchronous data. The higher the level of the data asynchronicity is, the slower the convergence rates are expected. As a major theoretical finding that confirms another merit of the proposed estimator, we show that the proposed localized estimator has the same accuracy as if the high-dimensional noises are directly observed in the sense of the same convergence rates.

The rest of this paper is organized as follows. The methodology is outlined in Section 2, followed by theoretical development in Section 3. Numerical examples with simulations and a real data analysis are presented in Section 4. We conclude the paper with some discussions in Section 5. All technical proofs are given in Section 6. In the end of this section, we introduce some notations. For a matrix $\mathbf{B} = (b_{i,j})_{s_1 \times s_2}$, let $|\mathbf{B}|_\infty = \max_{1 \leq i \leq s_1, 1 \leq j \leq s_2} |b_{i,j}|$, $\|\mathbf{B}\|_1 = \max_{1 \leq j \leq s_2} \sum_{i=1}^{s_1} |b_{i,j}|$, $\|\mathbf{B}\|_\infty = \max_{1 \leq i \leq s_1} \sum_{j=1}^{s_2} |b_{i,j}|$ and $\|\mathbf{B}\|_2 = \lambda_{\max}^{1/2}(\mathbf{B}\mathbf{B}^{\mathrm{T}})$ where $\lambda_{\max}(\mathbf{B}\mathbf{B}^{\mathrm{T}})$ denotes the largest eigenvalue of $\mathbf{B}\mathbf{B}^{\mathrm{T}}$. Denote by $\mathbb{I}(\cdot)$ the indicator function. For a countable set $\mathcal{G}$, we use $|\mathcal{G}|$ to denote its cardinality. For two sequences of positive numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exists a positive constant c such that $a_n/b_n \leq c$. We write $a_n \asymp b_n$ if and only if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously.

2 Methodology

2.1 Model and data

The setting of our study contains the signal part – a p-dimensional continuous-time process $(\mathbf{X}_t)_{t \in [0,T]}$, where $[0,T]$ is the time frame in which the high frequency data are observed. We assume that $\mathbf{X}_t = (X_{1,t}, \ldots, X_{p,t})^{\mathrm{T}}$ satisfies:

$$\mathrm{d}X_{i,t} = \mu_{i,t}\,\mathrm{d}t + \sigma_{i,t}\,\mathrm{d}B_{i,t} \quad \text{and} \quad \mathbb{E}(\mathrm{d}B_{i,t} \cdot \mathrm{d}B_{j,t}) = \rho_{i,j,t}\,\mathrm{d}t \quad (i,j = 1,\ldots,p), \qquad (2.1)$$

where each drift process $\mu_{i,t}$ is a locally bounded and progressively measurable process, the spot volatility process $\sigma_{i,t}$ is a positive and locally bounded Itô semimartingale, and $B_{1,t}, \ldots, B_{p,t}$ are univariate standard Brownian motions.

For each $i = 1, \ldots, p$, we use $\mathcal{G}_i = \{t_{i,1}, \ldots, t_{i,n_i}\}$ to denote the grid of time points where we observe the noisy data of the ith component process $X_{i,t}$, where $0 \leq t_{i,1} < \cdots < t_{i,n_i} \leq T$. The subject-specific set $\mathcal{G}_i$ reflects the asynchronous nature of the problem. For the special case with synchronous data, all $\mathcal{G}_i$'s are the same. However, $\mathcal{G}_i$'s are typically different in many practical high-frequency data. Let n be the number of different time points in $\cup_{i=1}^p \mathcal{G}_i$, and we denote the different time points in $\cup_{i=1}^p \mathcal{G}_i$ by $0 \leq t_1 < \cdots < t_n \leq T$. For any $i,j = 1, \ldots, p$, we define

$$n_{i,j} = |\mathcal{G}_i \cap \mathcal{G}_j|,$$

where $n_{i,j}$ evaluates how many time points t_k's at which we observe the noisy data of the ith and jth component processes $X_{i,t}$ and $X_{j,t}$ simultaneously. Clearly, $n_{i,i} = n_i$ for any $i = 1, \ldots, p$.

We consider that the actual observed data are contaminated by additive measurement errors

4

in the sense that
$$Y_{i,t_{i,k}} = X_{i,t_{i,k}} + U_{i,t_{i,k}} \quad (i = 1, \ldots, p;\ k = 1, \ldots, n_i).$$

Formally, we write
$$\mathbf{Y}_{t_k} = \mathbf{X}_{t_k} + \mathbf{U}_{t_k} \quad (k = 1, \ldots, n). \tag{2.2}$$

At each time point t_k, we only observe $\sum_{i=1}^{p} \mathbb{I}(t_k \in \mathcal{G}_i)$ components of $\mathbf{Y}_{t_k}$. As a conventional assumption for analyzing noisy high-frequency data, $\mathbf{U}_{t_k}$ is taken to be independent of $\mathbf{X}_{t_k}$; see, among others, Aït-Sahalia, Fan and Xiu (2010), Xiu (2010), Liu and Tang (2014), Chang et al. (2018), Xia and Zheng (2018), and the monograph Aït-Sahalia and Jacod (2014). Specifically, we make the following assumption:

Assumption 1. The noises $\{\mathbf{U}_{t_k}\}_{k=1}^{n}$ are independently and identically distributed with mean zero, and are independent of the process $(\mathbf{X}_t)_{t \in [0,T]}$.

Clearly, our target $\mathbf{U}_{t_k}$ is a latent vector. To estimate its covariance matrix, eliminating the impact due to the latent process $\mathbf{X}_t$ is required, which means that now $\mathbf{U}_{t_k}$ performs like 'signal' but $\mathbf{X}_{t_k}$ is 'noise'. Our strategy is to perform a dedicated localization: focusing on observations that are in a close enough neighborhood. Formally, for any $i, j = 1, \ldots, p$, we write $\mathcal{G}_i \cap \mathcal{G}_j = \{t_{i,j,1}, \ldots, t_{i,j,n_{i,j}}\}$ with $t_{i,j,1} < \cdots < t_{i,j,n_{i,j}}$. Given $\xi > 0$, for each $k = 1, \ldots, n_{i,j}$, we define
$$S_{i,j,k} = \left\{ t_{i,j,\ell} \in \mathcal{G}_i \cap \mathcal{G}_j : |t_{i,j,\ell} - t_{i,j,k}| \leq \xi \text{ and } \ell \neq k \right\}. \tag{2.3}$$

Let $N_{i,j,k} = |S_{i,j,k}|$ and $\Delta t_{i,j,k} = t_{i,j,k+1} - t_{i,j,k}$ for any $k = 1, \ldots, n_{i,j} - 1$. In this paper, we consider the scenario with T being fixed but $\max_{1 \leq i,j \leq p} \max_{1 \leq k \leq n_{i,j}-1} \Delta t_{i,j,k} \to 0$ as $n \to \infty$. Formally, we make the following assumption:

Assumption 2. (i) As $n \to \infty$, $\min_{1 \leq i,j \leq p} \min_{1 \leq k \leq n_{i,j}-1} \Delta t_{i,j,k} / \max_{1 \leq i,j \leq p} \max_{1 \leq k \leq n_{i,j}-1} \Delta t_{i,j,k}$ is uniformly bounded away from zero and infinity. (ii) As $n \to \infty$, each $n_{i,j} \to \infty$, and $\min_{1 \leq i,j \leq p} n_{i,j} / \max_{1 \leq i,j \leq p} n$ is uniformly bounded away from zero and infinity.

The first part of Assumption 2 is a standard setting for studying high-frequency data. The second part requires enough number of pairwise synchronous observations. This is a reasonable practical setting; see also Aït-Sahalia, Fan and Xiu (2010) for a pairwise approach for estimating the realized covariance matrix for $(\mathbf{X}_t)_{t \in [0,T]}$. Based on part (ii) of Assumption 2, we write
$$\min_{1 \leq i,j \leq p} n_{i,j} \asymp \max_{1 \leq i,j \leq p} n_{i,j} \asymp n_*, \tag{2.4}$$
where $n_* \to \infty$ as $n \to \infty$. The setting with Assumption 2 is broad and general. As we will show in Theorems 1–4, the convergence rates for the estimates of the covariance matrix $\text{Cov}(\mathbf{U}_{t_k})$ will depend on n_* instead of n. In the special case with synchronous observations, we have $n_{i,j} = n$ for any $i, j = 1, \ldots, n$ and then we can set $n_* = n$. Then all our results also apply to the setting with synchronous data.

5

2.2   Covariance matrix estimation of $\mathbf{U}_{t_k}$

Let $\mathbf{\Sigma_u} = \mathrm{Cov}(\mathbf{U}_{t_k}) = (\sigma_{\mathbf{u},i,j})_{p \times p}$. It follows from Assumption 1 that $2\mathbf{\Sigma_u} = \mathrm{Cov}(\mathbf{U}_{t_k} - \mathbf{U}_{t_\ell})$ for any $\ell \neq k$. Notice that $\mathbf{U}_{t_k} - \mathbf{U}_{t_\ell} = (\mathbf{Y}_{t_k} - \mathbf{Y}_{t_\ell}) - (\mathbf{X}_{t_k} - \mathbf{X}_{t_\ell})$ and each $(X_{i,t})_{t \in [0,T]}$ is a continuous-time and continuous-path stochastic process, $|X_{i,t+h} - X_{i,t}| \to 0$ almost surely as $h \to 0$. Thus, in a small neighborhood $\mathcal{N}$ of t_k, the difference between the high-frequency observations $\mathbf{Y}_{t_k}$ and $\mathbf{Y}_{t_\ell}$, for $t_\ell \in \mathcal{N}$, can be approximately viewed as $\mathbf{U}_{t_k} - \mathbf{U}_{t_\ell}$. This suggests that, for any $i, j = 1, \dots, p$, we can estimate $\sigma_{\mathbf{u},i,j}$ by

$$\hat{\sigma}_{\mathbf{u},i,j}(\xi) = \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (Y_{i,t_{i,j,\ell}} - Y_{i,t_{i,j,k}})(Y_{j,t_{i,j,\ell}} - Y_{j,t_{i,j,k}}), \qquad (2.5)$$

where $S_{i,j,k}$ is defined as (2.3). Here the appropriate localization

$$\frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (Y_{i,t_{i,j,\ell}} - Y_{i,t_{i,j,k}})(Y_{j,t_{i,j,\ell}} - Y_{j,t_{i,j,k}})$$

tries to eliminate the impact due to the latent processes $X_{i,t}$ and $X_{j,t}$, and the size of the local neighborhood $S_{i,j,k}$ is governed by ξ. This parameter can be viewed as a parameter controlling the trade-off between the bias and variance of $\hat{\sigma}_{\mathbf{u},i,j}(\xi)$: a small ξ results in small bias, but also results in small $N_{i,j,k}$'s so that its variance becomes larger. On the other hand, a larger ξ induces a lower variance, but comes at the price of a larger bias due to the contribution of the dynamics in $X_{i,t}$ and $X_{j,t}$. For a fixed T, (2.4) and Assumption 2 imply that

$$\min_{1 \leq i,j \leq p} \min_{1 \leq k \leq n_{i,j}-1} \Delta t_{i,j,k} \asymp \max_{1 \leq i,j \leq p} \max_{1 \leq k \leq n_{i,j}-1} \Delta t_{i,j,k} \asymp n_*^{-1}.$$

Hence, it holds that

$$\min_{1 \leq i,j \leq p} \min_{1 \leq k \leq n_{i,j}} N_{i,j,k} \asymp \max_{1 \leq i,j \leq p} \max_{1 \leq k \leq n_{i,j}} N_{i,j,k} \asymp n_* \xi.$$

Let

$$\widehat{\mathbf{\Sigma}}_{\mathbf{u}}(\xi) = \{\hat{\sigma}_{\mathbf{u},i,j}(\xi)\}_{p \times p} \qquad (2.6)$$

for $\hat{\sigma}_{\mathbf{u},i,j}(\xi)$ defined as (2.5). Theorem 1 in Section 3 shows that the elements of $\widehat{\mathbf{\Sigma}}_{\mathbf{u}}(\xi)$ are uniformly consistent to the corresponding elements of $\mathbf{\Sigma_u}$ with suitable selection of ξ, i.e.

$$\mathbb{E}\{|\widehat{\mathbf{\Sigma}}_{\mathbf{u}}(\xi) - \mathbf{\Sigma_u}|_\infty\} \lesssim \sqrt{\frac{\log p}{n_*}}.$$

Theorem 2 in Section 3 shows that the associated convergence rate $n_*^{-1/2} \log^{1/2} p$ is the minimax optimal rate in the maximum element-wise loss for the covariance matrix estimations of the high-dimensional noise $\mathbf{U}_{t_k}$ in high-frequency data. More importantly, we know that $n_*^{-1/2} \log^{1/2} p$ is also

6

the minimax optimal rate in the maximum element-wise loss for the covariance matrix estimations of $\mathbf{U}_{t_k}$ if we have observations of the noises, which indicates that our estimator shares some oracle property and the proposed localization actually makes the impact of the latent process $\mathbf{X}_t$ be negligible.

However, the aforementioned element-wise consistency and optimality do not imply their counterparts for the covariance matrix estimation with high-dimensional data. That is, the estimator $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}(\xi)$ may not be consistent to $\boldsymbol{\Sigma}_{\mathbf{u}}$ under the spectral norm when $p \gg n$. This is a well-known phenomenon in high-dimensional covariance matrix estimation; see, among other, Bickel and Levina (2008a). For high-dimensional covariance matrix estimations, one often resorts to some classes of the target with extra information. With the extra information, the consistency under the spectral norm and other properties associated with the covariance matrix estimations can be well established. In this paper, we focus on the following class – the sparse covariance matrices considered in Bickel and Levina (2008b):

$$\mathcal{H}(q, c_p, M) = \left\{ \boldsymbol{\Sigma} = (\sigma_{i,j})_{p \times p} : \sigma_{i,i} \leq M \text{ and } \sum_{j=1}^{p} |\sigma_{i,j}|^q \leq c_p \text{ for all } i \right\}, \tag{2.7}$$

where $q \in [0, 1)$ and $M > 0$ are two prescribed constants, and c_p may diverge with p. Here c_p can be viewed as a parameter that characterizes the sparsity of $\boldsymbol{\Sigma}$, i.e., if c_p is smaller, then $\boldsymbol{\Sigma}$ is more sparse. If $q = 0$, we have

$$\mathcal{H}(0, c_p, M) = \left\{ \boldsymbol{\Sigma} = (\sigma_{i,j})_{p \times p} : \sigma_{i,i} \leq M \text{ and } \sum_{j=1}^{p} \mathbb{I}(\sigma_{i,j} \neq 0) \leq c_p \text{ for all } i \right\},$$

where c_p evaluates the number of nonzero components in each row of $\boldsymbol{\Sigma}$.

For $\boldsymbol{\Sigma}_{\mathbf{u}} \in \mathcal{H}(q, c_p, M)$, we propose the following thresholding estimator based on the element-wise estimation $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}(\xi)$:

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\text{thre}}(\xi) = \left[\hat{\sigma}_{\mathbf{u},i,j}(\xi) \mathbb{I}\left\{ |\hat{\sigma}_{\mathbf{u},i,j}(\xi)| \geq \beta (n_*^{-1} \log p)^{1/2} \right\} \right]_{p \times p}, \tag{2.8}$$

where $\beta > 0$ is a tuning parameter for the thresholding level. Theorem 3 in Section 3 indicates that such defined thresholding estimator $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\text{thre}}(\xi)$ is consistent to $\boldsymbol{\Sigma}_{\mathbf{u}}$ under the spectral norm with suitable selections of ξ and β, i.e.

$$\mathbb{E}\left\{ \|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\text{thre}}(\xi) - \boldsymbol{\Sigma}_{\mathbf{u}}\|_2^2 \right\} \lesssim c_p^2 \left(\frac{\log p}{n_*} \right)^{1-q}.$$

Furthermore, Theorem 4 in Section 3 shows that $c_p(n_*^{-1} \log p)^{(1-q)/2}$ is the minimax optimal convergence rate with the spectral norm loss function for the covariance matrix estimations of the high-dimensional noise $\mathbf{U}_{t_k}$ in high-frequency data. Again, this rate is also the minimax optimal convergence rate in the spectral norm loss if we have observations of the noises.

3 Theoretical analysis

In this section, we establish the theoretical properties of the proposed estimators. We require the following three assumptions.

Assumption 3. Write $\mathbf{U}_{t_k} = (U_{1,t_k}, \ldots, U_{p,t_k})^{\mathrm{T}}$. There exist constants $K_1 > 0$ and $K_2 > 0$ such that $\mathbb{P}(|U_{i,t_k}| > u) \leq K_1 \exp(-K_2 u^2)$ for any $i = 1, \ldots, p$, $k = 1, \ldots, n$ and $u > 0$.

Assumption 4. There exist constants $K_3 > 0$, $K_4 > 0$ and $K_5 > 0$ such that (i) $\mathbb{E}(\exp[\theta\{\mu_{i,t}^2 - \mathbb{E}(\mu_{i,t}^2)\}]) \leq \exp(K_4 \theta^2)$ and $\mathbb{E}(\exp[\theta\{\sigma_{i,t}^2 - \mathbb{E}(\sigma_{i,t}^2)\}]) \leq \exp(K_4 \theta^2)$ for any $i = 1, \ldots, p$, $0 \leq t \leq T$ and $\theta \in (0, K_3]$; (ii) $\mathbb{E}(\mu_{i,t}^2) \leq K_5$ and $\mathbb{E}(\sigma_{i,t}^2) \leq K_5$ for any $i = 1, \ldots, p$ and $0 \leq t \leq T$.

Assumption 5. There exist constants $\gamma > 0$, $K_6 > 0$ and $K_7 > 0$ such that

$$\mathbb{P}\left(\sup_{0 \leq t \leq T} \sigma_{i,t} > u\right) \leq K_6 \exp(-K_7 u^\gamma)$$

for any $i = 1, \ldots, p$ and $u > 0$.

All assumptions are mild for studying high-dimensional covariance matrix estimations with high-frequency data. Assumption 3 requires that each component of $\mathbf{U}_{t_k}$ is sub-gaussian. Following Lemma 2.2 of Petrov (1995), we know part (i) of Assumption 4 holds if there exist two positive constants C_1 and C_2 such that $\mathbb{P}\{|\mu_{i,t}^2 - \mathbb{E}(\mu_{i,t}^2)| \geq u\} \leq C_1 \exp(-C_2 u)$ and $\mathbb{P}\{|\sigma_{i,t}^2 - \mathbb{E}(\sigma_{i,t}^2)| \geq u\} \leq C_1 \exp(-C_2 u)$ for any $i = 1, \ldots, p$, $0 \leq t \leq T$ and $u > 0$. Assumption 5 describes the behavior of the tail probability of $\sup_{0 \leq t \leq T} \sigma_{i,t}$. If the spot volatility process $\sigma_{i,t}$ is uniformly bounded over $i = 1, \ldots, p$ and $0 \leq t \leq T$, then we can select $\gamma = \infty$ in Assumption 5. Then we have the following result.

Theorem 1. *Let $\mathcal{P}_1$ denote the collections of models for $\{\mathbf{Y}_{t_k}\}_{k=1}^n$ such that $\mathbf{Y}_{t_k} = \mathbf{X}_{t_k} + \mathbf{U}_{t_k}$, where the noises $\{\mathbf{U}_{t_k}\}_{k=1}^n$ satisfy Assumptions 3, $\mathbf{X}_t = (X_{1,t}, \ldots, X_{p,t})^{\mathrm{T}}$ follows the continuous-time diffusion process model (2.1) with each $\mu_{i,t}$ and $\sigma_{i,t}$ satisfying Assumptions 4 and 5, and the grids of time points $\{\mathcal{G}_i\}_{i=1}^p$ satisfy Assumption 2. Let $\xi \asymp n_*^{-\kappa}$ for some $\kappa \in (1/2, 1]$, where n_* is specified in (2.4). Under Assumption 1, it holds that*

$$\sup_{\mathcal{P}_1} \mathbb{E}\{|\widehat{\mathbf{\Sigma}}_{\mathbf{u}}(\xi) - \mathbf{\Sigma}_{\mathbf{u}}|_\infty\} \lesssim \sqrt{\frac{\log p}{n_*}}$$

provided that $\log p = o\{n_*^{\tau(\kappa,\gamma)}\}$ with $\tau(\kappa,\gamma) = \min\{\kappa/5, (2\kappa - 1)\gamma/(\gamma + 4)\}$.

Theorem 1 gives the uniform convergence rate of $|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}|$ over $i, j = 1, \ldots, p$. If we select $\kappa = 1$, then $\tau(1, \gamma) = \min\{1/5, \gamma/(\gamma + 4)\}$ and the dimension of $\mathbf{U}_{t_k}$ can be as large as

$\exp[o\{n^{\tau(1,\gamma)}\}]$. In addition, if there is a positive constant C_3 such that $|\sigma_{i,t}| \leq C_3$ holds uniformly over $i = 1, \ldots, p$ and $0 \leq t \leq T$, then $\gamma = \infty$ and $\tau(1, \infty) = 1/5$. Furthermore, Theorem 2 below shows that the convergence rate $n_*^{-1/2} \log^{1/2} p$ is minimax optimal in the maximum element-wise loss for the covariance matrix estimations of the high-dimensional noise $\mathbf{U}_{t_k}$ in high-frequency data.

Theorem 2. *Let $n/n_* \lesssim p$. Denote by $\check{\mathcal{F}}$ the class of all measurable functionals of the data. Then*

$$\inf_{\widehat{\boldsymbol{\Sigma}} \in \check{\mathcal{F}}} \sup_{\mathcal{P}_1} \mathbb{E}\big(|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{\mathbf{u}}|_\infty\big) \gtrsim \sqrt{\frac{\log p}{n_*}},$$

where $\mathcal{P}_1$ is defined in Theorem 1.

To establish the lower bound stated in Theorem 2, we essentially focus on a model belonging to $\mathcal{P}_1$ with $\mu_{i,t} = 0$ and $\sigma_{i,t} = 0$ for any $t \in [0, T]$ and $i = 1, \ldots, p$. Let $\mathcal{C}_k = \{1 \leq i \leq p : t_k \in \mathcal{G}_i\}$ for any $t = 1, \ldots, n$. In this specific model, the latent process $\mathbf{X}_t = \mathbf{0}$ for any $t \in [0, T]$ and thus the data we observed is $\mathcal{Z} = \{\mathbf{U}_{t_1, \mathcal{C}_1}, \ldots, \mathbf{U}_{t_n, \mathcal{C}_n}\}$. Here $\mathbf{U}_{t_k, \mathcal{C}_k}$ denotes the subvector of $\mathbf{U}_{t_k}$ with components indexed by $\mathcal{C}_k$. Hence, $n_*^{-1/2} \log^{1/2} p$ is also the minimax optimal rate in the maximum element-wise loss for the covariance matrix estimations of $\mathbf{U}_{t_k}$ with data $\mathcal{Z} = \{\mathbf{U}_{t_1, \mathcal{C}_1}, \ldots, \mathbf{U}_{t_n, \mathcal{C}_n}\}$, which indicates that the estimator $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}(\xi)$ shares some oracle property and the proposed localization actually makes the impact of the latent process $\mathbf{X}_t$ be negligible.

Concerning the loss function under the spectral norm for the whole covariance matrix estimation, Theorems 3 and 4 below establish the minimax optimal convergence rate of the thresholding estimator $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\mathrm{thre}}(\xi)$ in (2.8).

Theorem 3. *Let $\mathcal{P}_2$ denote the collections of models for $\{\mathbf{Y}_{t_k}\}_{k=1}^n$ such that $\mathbf{Y}_{t_k} = \mathbf{X}_{t_k} + \mathbf{U}_{t_k}$, where the noises $\{\mathbf{U}_{t_k}\}_{k=1}^n$ satisfy Assumptions 3 and the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{u}} \in \mathcal{H}(q, c_p, M)$, $\mathbf{X}_t = (X_{1,t}, \ldots, X_{p,t})^{\mathrm{T}}$ follows the continuous-time diffusion process model (2.1) with each $\mu_{i,t}$ and $\sigma_{i,t}$ satisfying Assumptions 4 and 5, and the grids of time points $\{\mathcal{G}_i\}_{i=1}^p$ satisfy Assumption 2. Let $\xi \asymp n_*^{-\kappa}$ for some $\kappa \in (1/2, 1]$, where n_* is specified in (2.4). Under Assumption 1, it holds that*

$$\sup_{\mathcal{P}_2} \mathbb{E}\big\{\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\mathrm{thre}}(\xi) - \boldsymbol{\Sigma}_{\mathbf{u}}\|_2^2\big\} \lesssim c_p^2 \left(\frac{\log p}{n_*}\right)^{1-q}$$

provided that $\log p = o\{n_*^{\tau(\kappa,\gamma)}\}$ with $\tau(\kappa, \gamma) = \min\{\kappa/5, (2\kappa - 1)\gamma/(\gamma + 4)\}$.

Our result in the following Theorem 4 justifies that the convergence rate $c_p(n_*^{-1} \log p)^{(1-q)/2}$ is minimax optimal under the spectral norm loss function for the covariance matrix estimations of $\mathbf{U}_{t_k}$ with the sparsity structure (2.7). Again, this rate is also the minimax optimal rate in the spectral norm loss for the covariance matrix estimations of $\mathbf{U}_{t_k}$ with data $\mathcal{Z} = \{\mathbf{U}_{t_1, \mathcal{C}_1}, \ldots, \mathbf{U}_{t_n, \mathcal{C}_n}\}$.

Theorem 4. *Let $n/n_* \lesssim p$. Denote by $\check{\mathcal{F}}$ the class of all measurable functionals of the data. Then*

$$\inf_{\widehat{\boldsymbol{\Sigma}} \in \check{\mathcal{F}}} \sup_{\mathcal{P}_2} \mathbb{E}\big(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{\mathbf{u}}\|_2^2\big) \gtrsim c_p^2 \left(\frac{\log p}{n_*}\right)^{1-q}$$

provided that $c_p \lesssim n_*^{(1-q)/2}(\log p)^{-(3-q)/2}$.

In summary, we conclude that it is n_* – the effective sample size of the pairwise synchronous observations – determining the convergence rate of the covariance matrix estimation of the noise $\mathbf{U}_{t_k}$. Practically, n_* is expected to be smaller than n – the total number of observation times. Hence, the accuracy of the covariance matrix estimation is affected by the level of data asynchronicity – the more asynchronous the data are, the more difficult it is to estimate $\boldsymbol{\Sigma}_{\mathbf{u}}$. Another finding from our theoretical analysis is that although the noises $\{\mathbf{U}_{t_1, \mathcal{C}_1}, \ldots, \mathbf{U}_{t_n, \mathcal{C}_n}\}$ are not directly observable, the localized estimator has the same accuracy as the one when the noises $\{\mathbf{U}_{t_1, \mathcal{C}_1}, \ldots, \mathbf{U}_{t_n, \mathcal{C}_n}\}$ are observed in the sense of the same convergence rates for estimating $\boldsymbol{\Sigma}_{\mathbf{u}}$ with high-frequency data. From the practical perspective, it can be viewed as a bless from the high-frequency data with adequate amount of data information locally, so that the statistical properties of the noises can be accurately revealed.

4 Numerical studies

4.1 Simulations

We now demonstrate the finite-sample performance of the proposed estimator by simulations. We generated each continuous-time process $X_{i,t}$ $(i = 1, \ldots, p)$ from the Heston model:

$$\mathrm{d}X_{i,t} = \sigma_{i,t} \, \mathrm{d}B_{i,t}, \qquad \mathrm{d}\sigma_{i,t}^2 = \kappa_i(\bar{\sigma}_i^2 - \sigma_{i,t}^2) \, \mathrm{d}t + s_i \sigma_{i,t} \, \mathrm{d}W_{i,t},$$

where $B_{1,t}, \ldots, B_{p,t}$ and $W_{1,t}, \ldots, W_{p,t}$ are univariate standard Brownian motions, $\mathbb{E}(\mathrm{d}B_{i,t} \cdot \mathrm{d}W_{j,t}) = \delta_{i,j} \varsigma_i \, \mathrm{d}t$ and $\mathbb{E}(\mathrm{d}B_{i,t} \cdot \mathrm{d}B_{j,t}) = \rho_{i,j} \, \mathrm{d}t$ with $\delta_{i,j} = \mathbb{I}(i = j)$. We set $(\kappa_i, s_i, \bar{\sigma}_i^2, \varsigma_i) = (4, 0.3, 0.3^2, -0.3)$ for all $i = 1, \ldots, p$, and $(\rho_{i,j})_{p \times p} = \{\mathrm{diag}(\mathbf{A})\}^{-1/2} \mathbf{A}\mathbf{A}^{\mathsf{T}} \{\mathrm{diag}(\mathbf{A})\}^{-1/2}$, where $\mathbf{A} = (a_{i,j})_{p \times p}$ is a lower triangular matrix with $a_{i,j} = (-0.8)^{|i-j|}$ for $i \geq j$. The first observation of volatility process $\sigma_{i,t}^2$ is sampled from a Gamma distribution $\Gamma(\kappa_i \bar{\sigma}_i^2/s_i^2, s_i^2/2\kappa_i)$. The setting of the model parameters is similar that in Aït-Sahalia and Yu (2009), which reflects the practical financial data scenarios; see also Xiu (2010), Liu and Tang (2014) and Chang et al. (2018).

Both synchronous and asynchronous high-frequency data were considered in our numerical studies. Following the convention that a financial year typically has 252 active trading days, we

took $t \in [0, T]$ with $T = 1/252$ corresponding one trading day. Using the convention of 6.5 business hours in a trading day, we consider the per-second high-frequency data with potentially $60 \times 60 \times 6.5 = 23,400$ observations. Let $\tilde{t}_k = k/(252 \times 23400)$ ($k = 1, \ldots, 23400$) denoting the relative time that an observation was taken. We first generated the data by $\mathbf{Y}_{\tilde{t}_k} = \mathbf{X}_{\tilde{t}_k} + \mathbf{U}_{\tilde{t}_k}$ ($k = 1, \ldots, 23400$) with $(\mathbf{X}_t)_{t \in [0,T]}$ generated from the Heston model mentioned above and the noises $\mathbf{U}_{\tilde{t}_k}$'s generated independently from $N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u}})$ with $\boldsymbol{\Sigma}_{\mathbf{u}} = 0.005^2 \mathbf{R}$. We considered the following three models for $\mathbf{R}$ with different setting of the correlations:

(M1) $\mathbf{R} = (r_{i,j})_{p \times p}$ is a banded matrix with $r_{i,i} = 1$, $r_{i+1,i} = r_{i,i+1} = 0.6$, $r_{i+2,i} = r_{i,i+2} = 0.3$, and $r_{i,j} = 0$ for $|i - j| \geq 3$.

(M2) $\mathbf{R} = \widetilde{\mathbf{R}} + \{|\lambda_{\min}(\widetilde{\mathbf{R}})| + 0.05\}\mathbf{I}_p$, $\lambda_{\min}(\widetilde{\mathbf{R}})$ is the smallest eigenvalue of $\widetilde{\mathbf{R}}$, and $\widetilde{\mathbf{R}} = (\tilde{r}_{i,j})_{p \times p}$ where $\tilde{r}_{i,j} = w_{i,j} b_{i,j}$, $w_{i,j}$'s are independently generated from uniform distribution $U(0.4, 0.8)$, $b_{i,j}$'s are independently generated from Bernoulli distribution with successful probability 0.04. We then let

(M3) $\mathbf{R} = (r_{i,j})_{p \times p}$ with $r_{i,j} = 0.6^{|i-j|}$.

To mimic the synchronous scenario with different numbers of the within-day high-frequency observations, we took the observed data set being $\{\mathbf{Y}_{\tilde{t}_{k\Delta}}\}_{k=1}^{\lfloor 23400/\Delta \rfloor}$ for some given Δ, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. Clearly, larger Δ means fewer number of observations in the simulated data sets. Then the time points $\{t_1, \ldots, t_n\}$ where we observed the noisy data satisfy $t_k = \tilde{t}_{k\Delta}$ for each $k = 1, \ldots, \lfloor 23400/\Delta \rfloor$. For asynchronous scenario, we took a second step after generating the data $\{\mathbf{Y}_{\tilde{t}_k}\}_{k=1}^{23400}$ by applying p independent Poisson processes with intensity parameter λ to sequentially determine whether or not the $Y_{i,\tilde{t}_k}$ is observed at $\tilde{t}_k$. This implies that on average there are $\lfloor 23400/\lambda \rfloor$ observations for each component process. For asynchronous data, not all component processes are observed at the same time points, and we recall that $\{t_{i,j,1}, \ldots, t_{i,j,n_{i,j}}\}$ is the set of the time points at which we observed both the ith and jth component processes simultaneously.

The proposed estimator $\hat{\sigma}_{\mathbf{u},i,j}(\xi)$ in (2.5) involves a tuning parameter $\xi > 0$, so that selecting ξ is a relevant objective in practice. As in Theorems 1 and 3, to allow that dimension p diverges adequately fast, we need to select $\xi \asymp n_*^{-1}$, which means that the localization step of the proposed estimator only involves those data in a small neighborhood. With this principle in mind, we can use the following analogue in practice

$$\hat{\sigma}_{\mathbf{u},i,j} = \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}^*} \sum_{0 < |\ell - k| \leq K} (Y_{i,t_{i,j,\ell}} - Y_{i,t_{i,j,k}})(Y_{j,t_{i,j,\ell}} - Y_{j,t_{i,j,k}}) =: \frac{1}{n_{i,j}} \sum_{k=1}^{n_{i,j}} \zeta_{i,j,k}$$

11

for some small positive integer K, and $N^*_{i,j,k} = |\{t_{i,j,\ell} : 0 < |\ell - k| \leq K\}|$. Thus choosing K serves as the same role as choosing ξ. We attempted different settings with $K = 6, 7$ and 8 in our simulations. Based on $\hat{\sigma}_{\mathbf{u},i,j}$'s, we implemented the thresholding estimator as follows:

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\mathrm{thre}} = (\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}})_{p \times p} = \left[\hat{\sigma}_{\mathbf{u},i,j} \mathbb{I}\left\{ |\hat{\sigma}_{\mathbf{u},i,j}| \geq \varpi_{i,j} \right\} \right]_{p \times p} \tag{4.1}$$

where $\varpi_{ij} = 2\hat{\theta}_{i,j}^{1/2} n_{i,j}^{-1/2} \log^{1/2} p$ with $\hat{\theta}_{i,j}$ being an estimate of the long-run variance of the process $\{\zeta_{i,j,k}\}_{k=1}^{n_{i,j}}$. We chose $\hat{\theta}_{i,j}$ as the following kernel-type estimator:

$$\hat{\theta}_{i,j} = \sum_{\ell=-n_{i,j}+1}^{n_{i,j}-1} \mathcal{K}\left(\frac{\ell}{h}\right) \hat{H}_{i,j}(\ell)$$

where $\mathcal{K}(\cdot)$ is a kernel function, h is the bandwidth, $\hat{H}_{i,j}(\ell) = n_{i,j}^{-1} \sum_{k=\ell+1}^{n_{i,j}} (\zeta_{i,j,k} - \hat{\sigma}_{\mathbf{u},i,j} - \bar{\zeta}_{i,j})(\zeta_{i,j,k-\ell} - \hat{\sigma}_{\mathbf{u},i,j} - \bar{\zeta}_{i,j})$ if $\ell \geq 0$ and $\hat{H}_{i,j}(\ell) = n_{i,j}^{-1} \sum_{k=-\ell+1}^{n_{i,j}} (\zeta_{i,j,k+\ell} - \hat{\sigma}_{\mathbf{u},i,j} - \bar{\zeta}_{i,j})(\zeta_{i,j,k} - \hat{\sigma}_{\mathbf{u},i,j} - \bar{\zeta}_{i,j})$ otherwise, where $\bar{\zeta}_{i,j} = n_{i,j}^{-1} \sum_{k=1}^{n_{i,j}} (\zeta_{i,j,k} - \hat{\sigma}_{\mathbf{u},i,j})$. Andrews (1991) suggested to adopt the quadratic spectral kernel

$$\mathcal{K}(x) = \frac{25}{12\pi^2 x^2} \left\{ \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right\}$$

with optimal bandwidth

$$h = 1.3221\left\{ 4n_{i,j}\hat{\vartheta}_{i,j}^2 (1 - \hat{\vartheta}_{i,j})^{-4} \right\}^{1/5},$$

where $\hat{\vartheta}_{i,j}$ is the estimated autoregressive coefficient from fitting an AR(1) model to time series $\{\zeta_{i,j,k}\}_{k=1}^{n_{i,j}}$.

For each setting, we repeated the experiment 1000 times, and for each repetition we evaluated $\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\mathrm{thre}} - \boldsymbol{\Sigma}_{\mathbf{u}}\|_F / \|\boldsymbol{\Sigma}_{\mathbf{u}}\|_F$ measuring the relative estimation error. Here $\|\cdot\|_F$ denotes the matrix Frobenious norm. The average of the relative estimation errors with respect to the simulations for different settings are summarized in Table 1.

We have several observations. First, we find that in general, our estimator performed quite well with satisfactorily small relative estimation errors for all cases. Second, as the dimension p increases, the relative estimation errors worsen a bit, but at a very slow pace growing with p. This demonstrated the promising performance of the thresholding method for handling high-dimensional covariance estimations. Third, as the sampling frequency became higher (smaller Δ or λ), the performance is seen improved by observing smaller relative estimation errors, reflecting the blessing to the covariance estimations with more high-frequency data. This is actually the reason why the performance of the estimator with synchronous data is better than that of the asynchronous data when Δ and λ are the same. Fourth, we find that the differences are small among the performances with different tuning parameters $K = 6, 7$ and 8. We have also tried different values for K and find that the results are similar, especially when the effective sample

12

sizes are large. This suggests that the performance of the estimator is not sensitive to the choice of the tuning parameter.

Table 1: Averages of the relative estimation errors ($\times 100$) in different settings.

Synchronous		M1			M2			M3		
p	K	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$
	6	2.29	1.85	1.30	2.26	1.84	1.29	5.20	4.30	3.16
50	7	2.31	1.86	1.30	2.28	1.85	1.30	5.01	4.15	3.06
	8	2.34	1.88	1.31	2.33	1.88	1.31	4.86	4.03	2.97
	6	2.32	1.87	1.31	2.67	2.17	1.52	5.36	4.44	3.27
100	7	2.36	1.89	1.31	2.71	2.19	1.53	5.18	4.28	3.17
	8	2.43	1.93	1.33	2.78	2.24	1.55	5.03	4.17	3.07
	6	2.36	1.89	1.32	3.33	2.70	1.90	5.45	4.51	3.32
200	7	2.43	1.93	1.33	3.39	2.74	1.91	5.27	4.36	3.22
	8	2.55	2.01	1.36	3.50	2.81	1.94	5.15	4.26	3.13
Aynchronous		M1			M2			M3		
p	K	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$
	6	4.41	3.12	1.97	4.09	2.80	1.86	9.97	7.48	4.75
50	7	4.57	3.18	1.98	4.14	2.89	1.88	9.73	7.18	4.57
	8	4.75	3.27	2.01	4.31	3.01	1.92	9.60	6.97	4.42
	6	4.54	3.18	1.98	5.27	3.54	2.27	10.24	7.68	4.89
100	7	4.78	3.29	2.01	5.39	3.66	2.30	10.03	7.40	4.71
	8	5.07	3.45	2.06	5.64	3.86	2.37	9.94	7.21	4.57
	6	4.75	3.28	2.01	7.03	4.67	2.91	10.43	7.80	4.97
200	7	5.18	3.49	2.07	7.20	4.85	2.96	10.30	7.55	4.80
	8	5.66	3.76	2.18	7.52	5.11	3.06	10.31	7.42	4.68

In addition, we also calculated in Table 2 the true positive rate (TPR) and the false positive rate (FPR) with definition

$$\text{TPR} = \frac{|\{(i,j) : \hat{\sigma}^{\text{thre}}_{\mathbf{u},i,j} \neq 0 \text{ and } \sigma_{\mathbf{u},i,j} \neq 0\}|}{|\{(i,j) : \sigma_{\mathbf{u},i,j} \neq 0\}|},$$

$$\text{FPR} = \frac{|\{(i,j) : \hat{\sigma}^{\text{thre}}_{\mathbf{u},i,j} \neq 0 \text{ and } \sigma_{\mathbf{u},i,j} = 0\}|}{|\{(i,j) : \sigma_{\mathbf{u},i,j} = 0\}|}.$$

Since the covariance matrix considered in M3 has no exact zero elements, reporting the TPR and FPR for M3 is not sensible. Table 2 shows that the TPR for all cases are equal to 1 or quite close to 1, and the FPR for all cases are almost 0. This indicates that our proposed thresholding method can recover the zero elements of the covariance matrix very accurately.

13

Table 2: TPR (×100) and FPR (×100) in different settings.

| | | | TPR | | | | | | FPR | | | | | |
| | | | Synchronous (Δ) | | | Asynchronous (λ) | | | Synchronous (Δ) | | | Asynchronous (λ) | | |
	p	K	3	2	1	3	2	1	3	2	1	3	2	1
		6	100	100	100	100	100	100	0.01	0.01	0.01	0.05	0.03	0.01
	50	7	100	100	100	100	100	100	0.03	0.02	0.01	0.11	0.06	0.03
		8	100	100	100	100	100	100	0.07	0.05	0.03	0.20	0.12	0.06
		6	100	100	100	100	100	100	0.01	0.01	0.01	0.05	0.03	0.01
M1	100	7	100	100	100	100	100	100	0.03	0.02	0.01	0.11	0.06	0.03
		8	100	100	100	100	100	100	0.07	0.05	0.03	0.18	0.12	0.06
		6	100	100	100	100	100	100	0.01	0.01	0.01	0.05	0.03	0.01
	200	7	100	100	100	100	100	100	0.03	0.02	0.01	0.11	0.06	0.03
		8	100	100	100	100	100	100	0.07	0.05	0.03	0.18	0.12	0.06
		6	100	100	100	99.34	100	100	0.01	0.01	0.01	0.05	0.02	0.01
	50	7	100	100	100	99.61	100	100	0.03	0.02	0.01	0.10	0.06	0.03
		8	100	100	100	99.72	100	100	0.07	0.05	0.03	0.18	0.12	0.06
		6	100	100	100	99.40	100	100	0.01	0.01	0.01	0.05	0.02	0.01
M2	100	7	100	100	100	99.67	100	100	0.03	0.02	0.01	0.11	0.06	0.03
		8	100	100	100	99.79	100	100	0.06	0.05	0.03	0.18	0.12	0.06
		6	100	100	100	99.36	100	100	0.01	0.01	0.01	0.05	0.03	0.01
	200	7	100	100	100	99.63	100	100	0.03	0.02	0.01	0.11	0.06	0.03
		8	100	100	100	99.78	100	100	0.07	0.05	0.03	0.18	0.12	0.06

4.2 Real data analysis

In this section, we apply our method to perform covariance matrix estimation of the microstructure noise in a real financial data set available from the TAQ database. The data set contains the tick-by-tick observations of the Standard and Poors (S&P) 500 constituent stocks on two days – November 4 and 22 – in 2016. We choose these two days because the levels of the CBOE Volatility Index (VIX) – an overall measure of the market variation level – are quite different. The VIX of November 4, 2016 is 22.51, which is the largest VIX in November 2016. The VIX of November 22, 2016 is 12.41, which is the second smallest VIX in November 2016; the smallest VIX occurred on November 25, 2016, which was the first trading day after the Thanksgiving holiday so we do not investigate it. By examining the covariance matrix estimation of the noise contaminated in the real log-prices, we attempt to reveal some different features.

Similar to that in Fan, Furger and Xiu (2016), we used the Global Industry Classification Standard (GICS) codes to sort the companies in S&P 500. The code is 8-digits and each company has its unique code. Digits 1-2 of the code describe the company's sector; digits 3-4 describe the industry group; digits 5-6 describe the industry; digits 7-8 describe the sub-industry. Based on the GICS codes, there are 36, 27, 71, 84, 36, 58, 64, 65, 5, 28, and 26 companies respectively belonging to the 11 different sectors – Energy (E), Materials (M), Industrials (I), Consumer Discretionary (D), Consumer Staples (S), Health Care (H), Financial (F), Information Technology (T), Telecommunication Services (C), Utilities (U), and Real Estate (R) respectively. Since there are only 5 companies belonged to Telecommunication Services, we therefore put companies in the Information Technology and Telecommunication Services together and denoted them as 'T'. We report in Figure 1 the magnitudes of the elements in the estimated correlation matrices of the microstructure noise based on our proposed covariance matrix estimator $\widehat{\Sigma}_{\mathbf{u}}^{\text{thre}}$ as in (4.1) respectively for November 4 and 22, 2016 with $K = 6$ and 8. The blue blocks along the diagonal in Figure 1 denote the sector classification according to the digits 1-2 of GICS codes, with each block containing correlations between the companies in the same sector.

Some interesting findings are illustrated by Figure 1. First, the sparsity level of the estimated correlation matrices is generally high with many components being estimated as zero, indicating that the sparsity condition imposed on the covariance matrix of microstructure noise is reasonable. Second, the overall level of correlations are seen different in these two days. On November 4 when the VIX level was high, the overall correlation level between different components of the microstructure noise is also found to be higher, and denser than that on November 22. Third, the structures of the estimated correlation matrices have no significant difference for $K = 6$ and 8. This also indicates that our proposed estimation is not sensitive for the choice of K. Fourth, we see that the correlations within each sector are clear, especially for the Energy and Financial

sectors. In contrast, the correlations between different blocks are weak. This finding can also be considered as a support in some sense for the validity of the block diagonal structure imposed on the covariance matrix of the residuals in the factor model constructed in Fan, Furger and Xiu (2016) for analyzing the log-prices data of stocks.
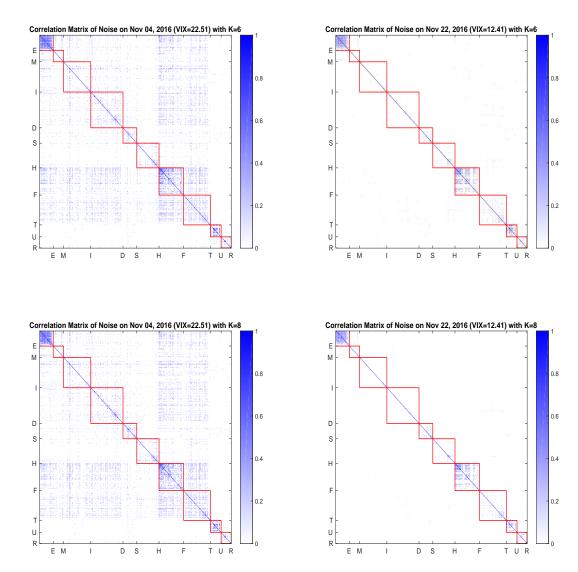
5   Discussion

In this paper, we consider estimating the covariance matrix of the high-dimensional noise in high-frequency data. We propose an estimator with appropriate localization and thresholding to achieve the minimax optimal convergence rates under two kinds of loss. Although all theoretical properties of the proposed estimator are derived under the continuous-time model (2.1), the method developed in this paper could be applied to other types of the process $\mathbf{X}_t$, such as the smooth ones typically encountered in the functional data literature. The key property that makes our method consistent is the continuity of the underlying process $\mathbf{X}_t$, but the convergence rate of the proposed estimator depends on more specific assumptions, such as those implied by the model (2.1). As the first attempt, we assume independent and identically distributed noises in our current setting. There are a few possibilities for extending the work to broad settings. First, we note that the problem becomes fundamentally different with new scopes when serial correlations are allowed for the noises. We anticipate different insights from such a setting on both the minimax optimality and the estimators, and we plan to address the problem in a comprehensive future study. Additionally, besides the sparsity setting where the thresholding estimator is appropriate, other approaches can also be developed attempting to use some structural information, for example, by employing a latent factor model. Then, the optimality, model specification, and testing problems are arising as interesting problems with new insights and challenges. We also plan to investigate those problems in future projects.

6   Proofs

In the sequel, we use C to denote a generic positive finite uniform constant that may be different in different uses.

Figure 1: Estimated correlation matrix of the microstructure noises on November 4 and 22, 2016

6.1 Proof of Theorem 1

By the definition of $\hat{\sigma}_{\mathbf{u},i,j}(\xi)$ as (2.5), we have

$$\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j} = \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (U_{i,t_{i,j,\ell}} - U_{i,t_{i,j,k}})(U_{j,t_{i,j,\ell}} - U_{j,t_{i,j,k}}) - \sigma_{\mathbf{u},i,j}}_{\mathrm{I}(i,j)}$$

$$+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}})(X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}})}_{\mathrm{II}(i,j)}$$

$$+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}})(U_{j,t_{i,j,\ell}} - U_{j,t_{i,j,k}})}_{\mathrm{III}(i,j)}$$

$$+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (U_{i,t_{i,j,\ell}} - U_{i,t_{i,j,k}})(X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}})}_{\mathrm{IV}(i,j)}.$$

Let $\xi \asymp n_*^{-\kappa}$ for $\kappa \in (1/2, 1]$. As we will show in Lemmas 1–3 that

$$\max_{1 \le i,j \le p} \mathbb{P}\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| > v\} \le C \exp(-Cn_* v^2) \tag{6.1}$$

for any $v = o(n_*^{\aleph})$ with $\aleph = \min\{(\kappa - 5)/10, (\kappa\gamma - \gamma - 2)/(\gamma + 4)\}$. Given sufficiently large $\alpha > 0$, it holds that

$$\mathbb{E}\{|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}(\xi) - \boldsymbol{\Sigma}_{\mathbf{u}}|_\infty\} \le \mathbb{E}\left[\max_{1 \le i,j \le p} |\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| \mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| \le \alpha\sqrt{\frac{\log p}{n_*}}\right\}\right]$$

$$+ \mathbb{E}\left[\max_{1 \le i,j \le p} |\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| \mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| > \alpha\sqrt{\frac{\log p}{n_*}}\right\}\right]$$

$$=: A_1 + A_2.$$

It is easy to see that $A_1 \le \alpha n_*^{-1/2} \log^{1/2} p$. By Cauchy-Schwarz inequality, we have

$$A_2 \le \sum_{i,j=1}^p \mathbb{E}\left[|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| \mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| > \alpha\sqrt{\frac{\log p}{n_*}}\right\}\right]$$

$$\le p^2 \max_{1 \le i,j \le p} \left[\mathbb{E}\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}|^2\}\right]^{1/2} \cdot \max_{1 \le i,j \le p} \left[\mathbb{P}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| > \alpha\sqrt{\frac{\log p}{n_*}}\right\}\right]^{1/2}.$$

18

If $\log p = o(n_*^{1+2\aleph})$, it follows from (6.1) that

$$\max_{1\le i,j\le p}\left[\mathbb{P}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}| > \alpha\sqrt{\frac{\log p}{n_*}}\right\}\right]^{1/2} \le Cp^{-w}$$

for sufficiently large $w > 0$. Here $w \to \infty$ as $\alpha \to \infty$. Following the proofs for Lemmas 1–3, we know

$$\max_{1\le i,j\le p}\left[\mathbb{E}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}|^2\right\}\right]^{1/2} \le C.$$

Therefore, $A_2 \le Cp^{2-w}$ which will be negligible in comparison to $n_*^{-1/2}\log^{1/2}p$ if we select α is sufficiently large. Then we have

$$\sup_{\mathcal{P}_1}\mathbb{E}\left\{|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}(\xi) - \boldsymbol{\Sigma}_{\mathbf{u}}|_\infty\right\} \le C\sqrt{\frac{\log p}{n_*}}.$$

We complete the proof of Theorem 1. $\qquad\square$

Lemma 1. *Under Assumptions 1–3, it holds that*

$$\max_{1\le i,j\le p}\mathbb{P}\{|I(i,j)| > v\} \le C\exp(-Cn_*v^2)$$

for any $v = o(n_*^{-1/2}\xi^{-1/10})$.

Proof. Recall that

$$I(i,j) = \underbrace{\frac{1}{n_{i,j}}\sum_{\ell=1}^{n_{i,j}}\left(\frac{1}{2} + \frac{1}{2}\sum_{t_{i,j,k}\in S_{i,j,\ell}}\frac{1}{N_{i,j,k}}\right)U_{i,t_{i,j,\ell}}U_{j,t_{i,j,\ell}} - \sigma_{\mathbf{u},i,j}}_{I_1(i,j)}$$
$$\underbrace{-\frac{1}{2n_{i,j}}\sum_{k=1}^{n_{i,j}}\frac{U_{i,t_{i,j,k}}}{N_{i,j,k}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}U_{j,t_{i,j,\ell}}}_{I_2(i,j)} \underbrace{-\frac{1}{2n_{i,j}}\sum_{k=1}^{n_{i,j}}\frac{U_{j,t_{i,j,k}}}{N_{i,j,k}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}U_{i,t_{i,j,\ell}}}_{I_3(i,j)}.$$

In the sequel, we will bound the tail probabilities of $I_1(i,j)$, $I_2(i,j)$ and $I_3(i,j)$, respectively.

Notice that $\sum_{\ell=1}^{n_{i,j}}\sum_{t_{i,j,k}\in S_{i,j,\ell}}N_{i,j,k}^{-1} = \sum_{k=1}^{n_{i,j}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}N_{i,j,k}^{-1} = n_{i,j}$, then we have

$$I_1(i,j) = \frac{1}{n_{i,j}}\sum_{\ell=1}^{n_{i,j}}\left(\frac{1}{2} + \frac{1}{2}\sum_{t_{i,j,k}\in S_{i,j,\ell}}\frac{1}{N_{i,j,k}}\right)(U_{i,t_{i,j,\ell}}U_{j,t_{i,j,\ell}} - \sigma_{\mathbf{u},i,j}) =: \frac{1}{n_{i,j}}\sum_{\ell=1}^{n_{i,j}}\zeta_{i,j,\ell}.$$

It follows from Assumption 2 that $\min_{1\le i,j\le p}\min_{1\le k\le n_{i,j}}N_{i,j,k} \asymp \max_{1\le i,j\le p}\max_{1\le k\le n_{i,j}}N_{i,j,k} \asymp n_*\xi$, which implies that $C^{-1} < \sum_{t_{i,j,k}\in S_{i,j,\ell}}N_{i,j,k}^{-1} < C$ holds uniformly over $\ell = 1,\ldots,n_{i,j}$ and

19

$i, j = 1, \ldots, p$ for some $C > 1$. By Assumption 3 and Lemma 2 of Chang, Tang and Wu (2013), we know that $\max_{1 \leq i,j \leq p, 1 \leq \ell \leq n_{i,j}} \mathbb{P}(|\zeta_{i,j,\ell}| > v) \leq C \exp(-Cv)$ for any $v > 0$. Then it follows from Theorem 3.1 of Saulis and Statulevičius (1991) that

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{I}_1(i,j)| > v\} \leq C \exp(-Cn_* v^2) \tag{6.2}$$

for any $v = o(n_*^{-1/3})$.

For any fixed $i, j = 1, \ldots, p$, let $\eta_{i,j,k} = 2^{-1}(\max_{1 \leq \ell \leq n_{i,j}} N_{i,j,\ell}^{1/2}) U_{i,t_{i,j,k}} N_{i,j,k}^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}$. Then

$$\mathrm{I}_2(i,j) = \frac{1}{n_{i,j} \max_{1 \leq \ell \leq n_{i,j}} N_{i,j,\ell}^{1/2}} \sum_{k=1}^{n_{i,j}} \eta_{i,j,k} \,,$$

which implies

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{I}_2(i,j)| > v\} = \max_{1 \leq i,j \leq p} \mathbb{P}\left(\left|\frac{1}{n_{i,j}} \sum_{k=1}^{n_{i,j}} \eta_{i,j,k}\right| > v \max_{1 \leq \ell \leq n_{i,j}} N_{i,j,\ell}^{1/2}\right). \tag{6.3}$$

It follows from Assumption 2 that

$$C^{-1} |U_{i,t_{i,j,k}}| \left|\frac{1}{N_{i,j,k}^{1/2}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right| < |\eta_{i,j,k}| < C |U_{i,t_{i,j,k}}| \left|\frac{1}{N_{i,j,k}^{1/2}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right|$$

holds uniformly over any $i, j = 1, \ldots, p$. By Lemma 1 of Chang, Tang and Wu (2013), we have that

$$\max_{1 \leq i,j \leq p, 1 \leq k \leq n_{i,j}} \mathbb{P}\left(\left|\frac{1}{N_{i,j,k}^{1/2}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right| \geq x\right)$$

$$\leq \begin{cases} C \exp(-Cx^2), & \text{if } 0 \leq x \leq C n_*^{1/2} \xi^{1/2}; \\ C \exp(-C n_*^{1/2} \xi^{1/2} x), & \text{if } x \geq C n_*^{1/2} \xi^{1/2}. \end{cases}$$

Therefore, for any $s \geq 2$, it holds that

$$\mathbb{E}\left(\left|\frac{1}{N_{i,j,k}^{1/2}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_\ell}\right|^s\right) = \int_0^\infty s x^{s-1} \mathbb{P}\left(\left|\frac{1}{N_{i,j,k}^{1/2}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right| \geq x\right) dx$$

$$\leq \frac{sC^{s-1}}{\sqrt{\pi}} \Gamma\left(\frac{s}{2}\right) + Cs(n_* \xi)^{s/2} \exp(-Cn_* \xi)$$

$$\leq C^s s^{(s+1)/2}.$$

Meanwhile, it holds that $\mathbb{E}(|U_{i,t_{i,j,k}}|^s) \leq C^s s^{(s+1)/2}$. By Cauchy-Schwarz inequality, we have

$$\mathbb{E}(|\eta_{i,j,k}|^s) \leq C\{\mathbb{E}(|U_{i,t_{i,j,k}}|^{2s})\}^{1/2} \left\{\mathbb{E}\left(\left|\frac{1}{N_{i,j,k}^{1/2}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right|^{2s}\right)\right\}^{1/2}$$

20

$$\leq C^s s^{(s+1)/2} \leq (s!)^2 C^s$$

for any $s \in \mathbb{N}$ and $s \geq 2$. Recall that $\{\eta_{i,j,k}\}_{k=1}^{n_{i,j}}$ is $(2\max_{1\leq\ell\leq n_{i,j}} N_{i,j,\ell})$-dependent. Let $W_{i,j}^2 = \mathbb{E}\{(\sum_{k=1}^{n_{i,j}} \eta_{i,j,k})^2\}$. It follows from Assumptions 1 and 2 that

$$
\frac{4W_{i,j}^2}{\max_{1\leq\ell\leq n_{i,j}} N_{i,j,\ell}} = \sum_{k=1}^{n_{i,j}} \mathbb{E}\left\{\left(\frac{U_{i,t_{i,j,k}}}{N_{i,j,k}} \sum_{t_{i,j,\ell}\in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right)^2\right\}
$$
$$
+ \sum_{k_1\neq k_2} \mathbb{E}\left\{\left(\frac{U_{i,t_{i,j,k_1}}}{N_{i,j,k_1}} \sum_{t_{i,j,\ell_1}\in S_{i,j,k_1}} U_{j,t_{i,j,\ell_1}}\right)\left(\frac{U_{i,t_{i,j,k_2}}}{N_{i,j,k_2}} \sum_{t_{i,j,\ell_2}\in S_{i,j,k_2}} U_{j,t_{i,j,\ell_2}}\right)\right\}
$$
$$
= \sigma_{\mathbf{u},i,i}\sigma_{\mathbf{u},j,j} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} + \sigma_{\mathbf{u},i,j}^2 \sum_{k=1}^{n_{i,j}} \sum_{t_{i,j,\ell}\in S_{i,j,k}} \frac{1}{N_{i,j,k}N_{i,j,\ell}}
$$
$$
\asymp \xi^{-1},
$$

which implies $W_{i,j} \asymp n_*^{1/2}$ holds uniformly over $i, j = 1, \ldots, p$. Theorem 4.30 of Saulis and Statulevičius (1991) implies that

$$
\max_{1\leq i,j\leq p} \mathbb{P}\left(\left|\frac{1}{W_{i,j}} \sum_{k=1}^{n_{i,j}} \eta_{i,j,k}\right| \geq x\right)
$$
$$
\leq
\begin{cases}
C\exp\left(-\frac{Cx^2}{n_*\xi}\right), & \text{if } 0 \leq x \leq Cn_*^{1/2}\xi^{2/5}; \\
C\exp\left\{-C\left(\frac{x}{n_*^{1/2}\xi}\right)^{1/3}\right\}, & \text{if } x \geq Cn_*^{1/2}\xi^{2/5}.
\end{cases}
$$

Thus, from (6.3), we have

$$
\max_{1\leq i,j\leq p} \mathbb{P}\{|\mathrm{I}_2(i,j)| > v\} \leq C\exp(-Cn_*v^2)
$$

for any $v = o(n_*^{-1/2}\xi^{-1/10})$. Similarly, we have

$$
\max_{1\leq i,j\leq p} \mathbb{P}\{|\mathrm{I}_3(i,j)| > v\} \leq C\exp(-Cn_*v^2)
$$

for any $v = o(n_*^{-1/2}\xi^{-1/10})$. Notice that $n_*\xi \geq C > 0$, then $n_*^{-1/2}\xi^{-1/10} \leq Cn_*^{-2/5} \leq Cn_*^{-1/3}$. Together with (6.2), we have

$$
\max_{1\leq i,j\leq p} \mathbb{P}\{|\mathrm{I}(i,j)| > v\} \leq C\exp(-Cn_*v^2)
$$

for any $v = o(n_*^{-1/2}\xi^{-1/10})$. We complete the proof of Lemma 1. $\qquad\square$

Lemma 2. *Let $\xi \asymp n_*^{-\kappa}$ for some $\kappa \in (1/2, 1]$. Under Assumptions 4 and 5, it holds that*

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{II}(i,j)| > v\} \leq C \exp(-C n_* v^2)$$

for any $v = o\{n_*^{(\kappa\gamma-\gamma-2)/(\gamma+4)}\}$, where γ is specified in Assumption 5.

Proof. Notice that $\mathrm{d}X_{i,t} = \mu_{i,t}\,\mathrm{d}t + \sigma_{i,t}\,\mathrm{d}B_{i,t}$. Then

$$\mathrm{II}(i,j) = \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s}\,\mathrm{d}s \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s}\,\mathrm{d}s \right)}_{\mathrm{II}_1(i,j)}$$

$$+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s}\,\mathrm{d}B_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s}\,\mathrm{d}B_{j,s} \right)}_{\mathrm{II}_2(i,j)}$$

$$+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s}\,\mathrm{d}s \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s}\,\mathrm{d}B_{j,s} \right)}_{\mathrm{II}_3(i,j)}$$

$$+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s}\,\mathrm{d}B_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s}\,\mathrm{d}s \right)}_{\mathrm{II}_4(i,j)} .$$

In the sequel, we will bound the tail probabilities of $\max_{1 \leq i,j \leq p} |\mathrm{II}_1(i,j)|$, $\max_{1 \leq i,j \leq p} |\mathrm{II}_2(i,j)|$, $\max_{1 \leq i,j \leq p} |\mathrm{II}_3(i,j)|$ and $\max_{1 \leq i,j \leq p} |\mathrm{II}_4(i,j)|$, respectively.

Let $\zeta_{i,j,k}^* = N_{i,j,k}^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s}\,\mathrm{d}s)(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s}\,\mathrm{d}s)$. Then $\mathrm{II}_1(i,j) = (2n_{i,j})^{-1} \sum_{k=1}^{n_{i,j}} \zeta_{i,j,k}^*$. We first bound $\mathbb{E}\{\exp(\theta\zeta_{i,j,k}^*)\}$ for any $|\theta| \in (0, K_3\xi^{-2}]$, where K_3 is specified in Assumption 4. By Jensen's inequality and Cauchy-Schwarz inequality,

$$\mathbb{E}\{\exp(\theta\zeta_{i,j,k}^*)\}$$

$$\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\left[\exp\left\{ \theta \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s}\,\mathrm{d}s \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s}\,\mathrm{d}s \right) \right\} \right]$$

$$\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \frac{1}{|t_{i,j,\ell} - t_{i,j,k}|^2}$$

$$\times \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \mathbb{E}\{\exp(\theta|t_{i,j,\ell} - t_{i,j,k}|^2 \mu_{i,s_1}\mu_{j,s_2})\}\,\mathrm{d}s_1\mathrm{d}s_2 \qquad (6.4)$$

$$\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \frac{1}{|t_{i,j,\ell} - t_{i,j,k}|^2} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \left[\mathbb{E}\{\exp(|\theta||t_{i,j,\ell} - t_{i,j,k}|^2 \mu_{i,s_1}^2)\} \right]^{1/2} \mathrm{d}s_1$$

22

$$\times \int_{t_{i,j,k}\wedge t_{i,j,\ell}}^{t_{i,j,k}\vee t_{i,j,\ell}} \left[\mathbb{E}\{\exp(|\theta||t_{i,j,\ell}-t_{i,j,k}|^2\mu_{j,s_2}^2)\}\right]^{1/2}ds_2$$

$$\leq \max_{1\leq i\leq p}\sup_{0\leq s\leq T}\mathbb{E}\{\exp(|\theta|\xi^2\mu_{i,s}^2)\}.$$

It follows from Assumption 4 that

$$\max_{1\leq i\leq p}\sup_{0\leq s\leq T}\mathbb{E}\{\exp(|\theta|\xi^2\mu_{i,s}^2)\}\leq \exp(K_3K_5)\exp(K_4\xi^4\theta^2)\leq C\exp(C\xi^{4-\tau_1}\theta^2)$$

for any $|\theta|\in(0,K_3\xi^{-2}]$ and $\tau_1\in(0,4)$. From (6.4), we have $\mathbb{E}\{\exp(\theta\zeta_{i,j,k}^*)\}\leq C\exp(C\xi^{4-\tau_1}\theta^2)$ for any $|\theta|\in(0,K_3\xi^{-2}]$. By Lemma 2 of Fan, Li and Yu (2012), it holds that

$$\max_{1\leq i,j\leq p}\mathbb{P}\{|\mathrm{II}_1(i,j)|>v\}\leq C\exp(-Cv^2\xi^{\tau_1-4}) \tag{6.5}$$

for any $v=o(\xi^{2-\tau_1})$.

Let $\zeta_{i,j,k}^{**}=N_{i,j,k}^{-1}\sum_{t_{i,j,\ell}\in S_{i,j,k}}(\int_{t_{i,j,k}}^{t_{i,j,\ell}}\sigma_{i,s}dB_{i,s})(\int_{t_{i,j,k}}^{t_{i,j,\ell}}\sigma_{j,s}dB_{j,s})$. Then $\mathrm{II}_2(i,j)=(2n_{i,j})^{-1}\sum_{k=1}^{n_{i,j}}\zeta_{i,j,k}^{**}$. For any constant $d\in(0,\xi^{-1/2}]$, define a stopping time $\Gamma_{i,d}=\inf\{t:\sup_{0\leq s\leq t}\sigma_{i,s}>d\}\wedge T$. For any $|\theta|\in(0,d^{-2}\xi^{-1}/4]$, by Jensen's inequality and Cauchy-Schwarz inequality,

$$\mathbb{E}\{\exp(\theta\zeta_{i,j,k}^{**})\mathbb{I}(\Gamma_{i,d}=\Gamma_{j,d}=T)\}$$

$$\leq \frac{1}{N_{i,j,k}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}\mathbb{E}\left[\exp\left\{\theta\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}}\sigma_{i,s}dB_{i,s}\right)\right.\right.$$

$$\left.\left.\times\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}}\sigma_{j,s}dB_{j,s}\right)\right\}\mathbb{I}(\Gamma_{i,d}=\Gamma_{j,d}=T)\right] \tag{6.6}$$

$$\leq \frac{1}{N_{i,j,k}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}\left(\mathbb{E}\left[\exp\left\{|\theta|\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}}\sigma_{i,s}dB_{i,s}\right)^2\right\}\mathbb{I}(\Gamma_{i,d}=T)\right]\right)^{1/2}$$

$$\times\left(\mathbb{E}\left[\exp\left\{|\theta|\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}}\sigma_{j,s}dB_{j,s}\right)^2\right\}\mathbb{I}(\Gamma_{j,d}=T)\right]\right)^{1/2},$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Restricted on the event $\{\Gamma_{i,d}=T\}$, we have $\sup_{0\leq s\leq T}\sigma_{i,s}\leq d$. Then for any $|\theta|\in(0,d^{-2}\xi^{-1}/4]$ it holds that

$$\exp\left\{|\theta|\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}}\sigma_{i,s}dB_{i,s}\right)^2\right\}\mathbb{I}(\Gamma_{i,d}=T)$$

$$\leq \exp\left[|\theta|\left\{\left(\int_{t_{i,j,\ell}\wedge t_{i,j,k}}^{t_{i,j,\ell}\vee t_{i,j,k}}\sigma_{i,s}dB_{i,s}\right)^2-\int_{t_{i,j,\ell}\wedge t_{i,j,k}}^{t_{i,j,\ell}\vee t_{i,j,k}}\sigma_{i,s}^2ds\right\}\right]\mathbb{I}(\Gamma_{i,d}=T)$$

$$\times\exp\left(|\theta|\int_{t_{i,j,\ell}\wedge t_{i,j,k}}^{t_{i,j,\ell}\vee t_{i,j,k}}\sigma_{i,s}^2ds\right)\mathbb{I}(\Gamma_{i,d}=T) \tag{6.7}$$

$$\leq C\exp\left[|\theta|\left\{\left(\int_{t_{i,j,\ell}\wedge t_{i,j,k}}^{t_{i,j,\ell}\vee t_{i,j,k}}\sigma_{i,s}dB_{i,s}\right)^2-\int_{t_{i,j,\ell}\wedge t_{i,j,k}}^{t_{i,j,\ell}\vee t_{i,j,k}}\sigma_{i,s}^2ds\right\}\right]\mathbb{I}(\Gamma_{i,d}=T).$$

23

Recall $d \leq \xi^{-1/2}$. Following the arguments of Equation (A.5) in Fan, Li and Yu (2012), we have that

$$\mathbb{E}\left\{\exp\left[|\theta|\left\{\left(\int_{t_{i,j,\ell}\wedge t_{i,j,k}}^{t_{i,j,\ell}\vee t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right)^2 - \int_{t_{i,j,\ell}\wedge t_{i,j,k}}^{t_{i,j,\ell}\vee t_{i,j,k}}\sigma_{i,s}^2\,\mathrm{d}s\right\}\right]\mathbb{I}(\Gamma_{i,d}=T)\,\bigg|\,\mathcal{F}_{i,t_k}\right\}$$

$$\leq \mathbb{E}[\exp\{|\theta|(B_{i,d^2|t_{i,j,\ell}-t_{i,j,k}|}^2 - d^2|t_{i,j,\ell}-t_{i,j,k}|)\}] \tag{6.8}$$

$$= \mathbb{E}[\exp\{|\theta|d^2|t_{i,j,\ell}-t_{i,j,k}|(Z^2-1)\}]$$

$$\leq \exp(Cd^4\xi^2\theta^2) \leq \exp\{C(d\xi^{1/2})^{4-\tau_2}\theta^2\}$$

for any $\tau_2 \in (0,4)$, where $Z \sim N(0,1)$, and $\mathcal{F}_{i,t}$ denotes the σ-field generated by $(\sigma_{i,s}, B_{i,s})_{0\leq s\leq t}$. Therefore, by (6.6), we have

$$\mathbb{E}\{\exp(\theta\zeta_{i,j,k}^{**})\mathbb{I}(\Gamma_{i,d}=\Gamma_{j,d}=T)\} \leq C\exp\{C(d\xi^{1/2})^{4-\tau_2}\theta^2\}$$

for any $|\theta| \in (0, d^{-2}\xi^{-1}/4]$. By Lemma 2 of Fan, Li and Yu (2012), it holds that

$$\max_{1\leq i,j\leq p}\mathbb{P}\{|\mathrm{II}_2(i,j)| > v, \Gamma_{i,d}=\Gamma_{j,d}=T\} \leq C\exp\{-Cv^2(d\xi^{1/2})^{\tau_2-4}\}$$

for any $v = o\{(d\xi^{1/2})^{2-\tau_2}\}$. Notice that $\mathbb{P}\{|\mathrm{II}_2(i,j)| > v\} \leq \mathbb{P}\{|\mathrm{II}_2(i,j)| > v, \Gamma_{i,d} = \Gamma_{j,d} = T\} + \mathbb{P}(\Gamma_{i,d} \neq T) + \mathbb{P}(\Gamma_{j,d} \neq T)$. Since $\Gamma_{i,d} = \inf\{t : \sup_{0\leq s\leq t}\sigma_{i,s} > d\} \wedge T$, by Assumption 5, we have

$$\max_{1\leq i\leq p}\mathbb{P}(\Gamma_{i,d} \neq T) \leq \max_{1\leq i\leq p}\mathbb{P}\left(\sup_{0\leq s\leq T}\sigma_{i,s} > d\right) \leq C\exp(-Cd^\gamma).$$

Then

$$\max_{1\leq i,j\leq p}\mathbb{P}\{|\mathrm{II}_2(i,j)| > v\} \leq C\exp\{-Cv^2(d\xi^{1/2})^{\tau_2-4}\} + C\exp(-Cd^\gamma) \tag{6.9}$$

for any $v = o\{(d\xi^{1/2})^{2-\tau_2}\}$ with $d \leq \xi^{-1/2}$.

On the other hand, notice that $2|\mathrm{II}_3(i,j)| \leq |\mathrm{II}_1(i,i)| + |\mathrm{II}_2(j,j)|$ and $2|\mathrm{II}_4(i,j)| \leq |\mathrm{II}_1(j,j)| + |\mathrm{II}_2(i,i)|$, then (6.5) and (6.9) imply that

$$\max_{1\leq i,j\leq p}\mathbb{P}\{|\mathrm{II}_3(i,j)| > v\} \leq C\exp(-Cv^2\xi^{\tau_1-4})$$

$$+ C\exp\{-Cv^2(d\xi^{1/2})^{\tau_2-4}\} + C\exp(-Cd^\gamma)$$

and

$$\max_{1\leq i,j\leq p}\mathbb{P}\{|\mathrm{II}_4(i,j)| > v\} \leq C\exp(-Cv^2\xi^{\tau_1-4})$$

$$+ C\exp\{-Cv^2(d\xi^{1/2})^{\tau_2-4}\} + C\exp(-Cd^\gamma)$$

for any $v = o\{(d\xi^{1/2})^{2-\tau_2} \wedge \xi^{2-\tau_1}\}$ with $d \leq \xi^{-1/2}$. Together with (6.5) and (6.9), we have

$$\max_{1\leq i,j\leq p}\mathbb{P}\{|\mathrm{II}(i,j)| > v\} \leq C\exp(-Cv^2\xi^{\tau_1-4})$$

$$+ C\exp\{-Cv^2(d\xi^{1/2})^{\tau_2-4}\} + C\exp(-Cd^\gamma)$$

24

for any $v = o\{(d\xi^{1/2})^{2-\tau_2} \wedge \xi^{2-\tau_1}\}$ with $d \leq \xi^{-1/2}$.

Let $\xi \asymp n_*^{-\kappa}$ and $d \asymp n_*^{\vartheta}$ with $\kappa \in (1/2, 1]$ and $\vartheta \in (0, (2\kappa - 1)/4)$. If we require $\kappa(4 - \tau_1) \geq 1$ and $(\kappa/2 - \vartheta)(4 - \tau_2) \geq 1$, then

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{II}(i,j)| > v\} \leq C \exp(-Cn_* v^2) + C \exp(-Cn_*^{\vartheta\gamma})$$

for any $v = o(n_*^{\aleph_1})$ with $\aleph_1 = \min\{(\kappa/2 - \vartheta)(\tau_2 - 2), \kappa(\tau_1 - 2)\}$. Notice that $\kappa\tau_1 \leq 4\kappa - 1$ and $(\kappa/2 - \vartheta)\tau_2 \leq 2\kappa - 4\vartheta - 1$. To make $\aleph_1$ be largest, we select (τ_1, τ_2) satisfying $\kappa\tau_1 = 4\kappa - 1$ and $(\kappa/2 - \vartheta)\tau_2 = 2\kappa - 4\vartheta - 1$. Then $\aleph_1 = \kappa - 2\vartheta - 1$. Selecting $\vartheta = (2\kappa - 1)/(\gamma + 4)$, we have $\vartheta\gamma = 2\kappa - 4\vartheta - 1 \geq 2\aleph_1 + 1$. Therefore,

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{II}(i,j)| > v\} \leq C \exp(-Cn_* v^2)$$

for any $v = o\{n_*^{(\gamma\kappa - \gamma - 2)/(\gamma + 4)}\}$. We complete the proof of Lemma 2. $\qquad\square$

Lemma 3. *Let $\xi \asymp n_*^{-\kappa}$ for some $\kappa \in (0, 1]$. Under Assumptions 1–5, it holds that*

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{III}(i,j)| > v\} \leq C \exp(-Cn_* v^2)$$

and

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{IV}(i,j)| > v\} \leq C \exp(-Cn_* v^2)$$

for any $v = o\{n^{(\kappa\gamma - \gamma - 2)/(2\gamma + 4)}\}$, where γ is specified in Assumption 5.

Proof. By the definition of $\mathrm{III}(i,j)$, we can reformulate it as

$$\mathrm{III}(i,j) = \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}}\right)(X_{i,t_{i,j,k}} - X_{i,t_{i,j,\ell}})U_{j,t_{i,j,k}} .$$

For each $i, j = 1, \ldots, p$, define

$$D_{i,j} = \max_{1 \leq k \leq n_{i,j}} \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}}\right)(X_{i,t_{i,j,k}} - X_{i,t_{i,j,\ell}}) \right|$$

and

$$Q_{i,j}^2 = \sum_{k=1}^{n_{i,j}} \left\{ \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}}\right)(X_{i,t_{i,j,k}} - X_{i,t_{i,j,\ell}}) \right\}^2 .$$

We will first consider the tail probabilities $\mathbb{P}(D_{i,j} > v)$ and $\mathbb{P}(Q_{i,j}^2 > v)$.

By Bonferroni inequality, we have

$$\mathbb{P}(D_{i,j} > v) \leq \sum_{k=1}^{n_{i,j}} \mathbb{P}\left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}}\right)(X_{i,t_{i,j,k}} - X_{i,t_{i,j,\ell}}) \right| > v \right\}$$

25

$$\leq \sum_{k=1}^{n_{i,j}} \mathbb{P}\left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s}\, \mathrm{d}s \right| > \frac{v}{2} \right\} \tag{6.10}$$

$$+ \sum_{k=1}^{n_{i,j}} \mathbb{P}\left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s}\, \mathrm{d}B_{i,s} \right| > \frac{v}{2} \right\}.$$

For any $\theta > 0$, by Triangle inequality and Jensen's inequality, it holds that

$$\exp\left\{ \theta \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s}\, \mathrm{d}s \right| \right\}$$

$$\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \exp\left\{ \left(\frac{N_{i,j,k}}{N_{i,j,\ell}} + 1 \right) \int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \theta |\mu_{i,s}|\, \mathrm{d}s \right\}$$

$$\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \frac{1}{|t_{i,j,\ell} - t_{i,j,k}|} \int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \exp\left\{ \left(\frac{N_{i,j,k}}{N_{i,j,\ell}} + 1 \right) \theta |t_{i,j,\ell} - t_{i,j,k}| |\mu_{i,s}| \right\} \mathrm{d}s\,,$$

which implies that

$$\max_{1 \leq k \leq n_{i,j}} \mathbb{E}\left[\exp\left\{ \theta \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s}\, \mathrm{d}s \right| \right\} \right] \leq \sup_{0 \leq t \leq T} \mathbb{E}\{\exp(C\theta\xi|\mu_{i,t}|)\}\,.$$

It follows from Assumption 4 that

$$\sup_{0 \leq t \leq T} \mathbb{E}\{\exp(C\theta\xi|\mu_{i,t}|)\} \leq \sup_{0 \leq t \leq T} \mathbb{E}\left\{ \exp\left(K_3 |\mu_{i,t}|^2 + \frac{C^2 \xi^2 \theta^2}{4K_3} \right) \right\} \leq C \exp(C\xi^2\theta^2)\,.$$

Selecting $\theta \asymp \xi^{-1}$ and applying Markov's inequality, we have

$$\max_{1 \leq k \leq n_{i,j}} \mathbb{P}\left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s}\, \mathrm{d}s \right| > \frac{v}{2} \right\}$$

$$\leq \exp(-\theta v) \max_{1 \leq k \leq n_{i,j}} \mathbb{E}\left[\exp\left\{ 2\theta \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s}\, \mathrm{d}s \right| \right\} \right] \tag{6.11}$$

$$\leq C \exp(-C\xi^{-1} v)$$

for any $v > 0$.

For any constant $d \in (0, \xi^{-1/2}]$, define a stopping time $\Gamma_{i,d} = \inf\{t : \sup_{0 \leq s \leq t} \sigma_{i,s} > d\} \wedge T$. On the other hand, by Cauchy-Schwarz inequality, it holds that

$$\left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s}\, \mathrm{d}B_{i,s} \right|^2 \leq \frac{C}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s}\, \mathrm{d}B_{i,s} \right)^2,$$

which implies that

$$\mathbb{P}\left\{\left|\sum_{t_{i,j,\ell}\in S_{i,j,k}}\left(\frac{1}{N_{i,j,\ell}}+\frac{1}{N_{i,j,k}}\right)\int_{t_{i,j,\ell}}^{t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right|>\frac{v}{2},\Gamma_{i,d}=T\right\}$$

$$\leq\exp(-C\theta v^2)\mathbb{E}\left[\exp\left\{\frac{\theta}{N_{i,j,k}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}\left(\int_{t_{i,j,\ell}}^{t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right)^2\right\}\mathbb{I}(\Gamma_{i,d}=T)\right] \tag{6.12}$$

for any $\theta>0$. By Jensen's inequality, we have

$$\mathbb{E}\left[\exp\left\{\frac{\theta}{N_{i,j,k}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}\left(\int_{t_{i,j,\ell}}^{t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right)^2\right\}\mathbb{I}(\Gamma_{i,d}=T)\right]$$

$$\leq\frac{1}{N_{i,j,k}}\sum_{t_{i,j,\ell}\in S_{i,j,k}}\mathbb{E}\left[\exp\left\{\theta\left(\int_{t_{i,j,\ell}}^{t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right)^2\right\}\mathbb{I}(\Gamma_{i,d}=T)\right].$$

Same as (6.7) and (6.8), for any $\theta\in(0,d^{-2}\xi^{-1}/4]$, it holds that

$$\mathbb{E}\left[\exp\left\{\theta\left(\int_{t_{i,j,\ell}}^{t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right)^2\right\}\mathbb{I}(\Gamma_{i,d}=T)\right]\leq\exp(Cd^4\xi^2\theta^2)\leq C.$$

Selecting $\theta=d^{-2}\xi^{-1}/4$, together with (6.12), we have

$$\max_{1\leq k\leq n_{i,j}}\mathbb{P}\left\{\left|\sum_{t_{i,j,\ell}\in S_{i,j,k}}\left(\frac{1}{N_{i,j,\ell}}+\frac{1}{N_{i,j,k}}\right)\int_{t_\ell}^{t_k}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right|>\frac{v}{2},\Gamma_{i,d}=T\right\}\leq C\exp(-Cd^{-2}\xi^{-1}v^2)$$

for any $v>0$. It follows from Assumption 5 that

$$\max_{1\leq k\leq n_{i,j}}\mathbb{P}\left\{\left|\sum_{t_{i,j,\ell}\in S_{i,j,k}}\left(\frac{1}{N_{i,j,\ell}}+\frac{1}{N_{i,j,k}}\right)\int_{t_{i,j,\ell}}^{t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right|>\frac{v}{2}\right\}$$

$$\leq\mathbb{P}(\Gamma_{i,d}\neq T)+\max_{1\leq k\leq n_{i,j}}\mathbb{P}\left\{\left|\sum_{t_{i,j,\ell}\in S_{i,j,k}}\left(\frac{1}{N_{i,j,\ell}}+\frac{1}{N_{i,j,k}}\right)\int_{t_{i,j,\ell}}^{t_{i,j,k}}\sigma_{i,s}\,\mathrm{d}B_{i,s}\right|>\frac{v}{2},\Gamma_{i,d}=T\right\}$$

$$\leq C\exp(-Cd^\gamma)+C\exp(-Cd^{-2}\xi^{-1}v^2)$$

for any $v>0$. Letting $d\to\infty$, together with (6.11), (6.10) implies that

$$\max_{1\leq i,j\leq p}\mathbb{P}(D_{i,j}>v)\leq Cn_*\exp(-Cd^\gamma)+Cn_*\exp(-Cd^{-2}\xi^{-1}v^2) \tag{6.13}$$

for any $0<v\leq C$. Notice that $n_{i,j}D_{i,j}^2\geq Q_{i,j}^2$. Then (6.13) implies that

$$\max_{1\leq i,j\leq p}\mathbb{P}(Q_{i,j}^2>v)\leq\max_{1\leq i,j\leq p}\mathbb{P}(D_{i,j}>n_{i,j}^{-1/2}v^{1/2})$$

$$\leq Cn_*\exp(-Cd^\gamma)+Cn_*\exp(-Cd^{-2}\xi^{-1}n_*^{-1}v) \tag{6.14}$$

27

for any $0 < v \leq Cn_*$.

By Theorem 3.1 of Saulis and Statulevičius (1991), we have

$$\mathbb{P}\{|\mathrm{III}(i,j)| > x \,|\, (\mathbf{X}_t)_{t \in [0,T]}\} \leq \begin{cases} C \exp\left(- \dfrac{Cn_*^2 x^2}{Q_{i,j}^2} \right), & \text{if } 0 \leq x < \dfrac{CQ_{i,j}^2}{n_* D_{i,j}}; \\[4mm] C \exp\left(- \dfrac{Cn_* x}{D_{i,j}} \right), & \text{if } x \geq \dfrac{CQ_{i,j}^2}{n_* D_{i,j}}; \end{cases}$$

which implies that

$$\mathbb{P}\{|\mathrm{III}(i,j)| > x\} \leq C\mathbb{E}\left\{ \exp\left(- \frac{Cn_*^2 x^2}{Q_{i,j}^2} \right) \right\} + C\mathbb{E}\left\{ \exp\left(- \frac{Cn_* x}{D_{i,j}} \right) \right\}.$$

It follows from (6.13) with $v \asymp 1$ that

$$\mathbb{E}\left\{ \exp\left(- \frac{Cn_* x}{D_{i,j}} \right) \right\} \leq \exp(-Cv^{-1} n_* x) + \mathbb{P}(D_{i,j} > v)$$

$$\leq \exp(-Cn_* x) + Cn_* \exp(-Cd^\gamma) + Cn_* \exp(-Cd^{-2}\xi^{-1}).$$

Meanwhile, it follows from (6.14) with $v \asymp n_*^\delta$ for $\delta \in (0,1]$ that

$$\mathbb{E}\left\{ \exp\left(- \frac{Cn_*^2 x^2}{Q_{i,j}^2} \right) \right\} \leq \exp(-Cv^{-1} n_*^2 x^2) + \mathbb{P}(Q_{i,j}^2 > v)$$

$$\leq \exp(-Cn_*^{2-\delta} x^2) + Cn_* \exp(-Cd^\gamma) + Cn_* \exp(-Cd^{-2}\xi^{-1} n_*^{-1+\delta}).$$

Hence, it holds that

$$\mathbb{P}\{|\mathrm{III}(i,j)| > x\} \leq \exp(-Cn_* x) + \exp(-Cn_*^{2-\delta} x^2)$$
$$+ Cn_* \exp(-Cd^\gamma) + Cn_* \exp(-Cd^{-2}\xi^{-1} n_*^{-1+\delta})$$

for any $x > 0$. Let $\xi \asymp n_*^{-\kappa}$ for $\kappa \in (1-\delta, 1]$ and $d \asymp n_*^\vartheta$ for $\vartheta \in (0, (\kappa - 1 + \delta)/2)$. If we require $0 < x \leq C$, it holds that $n_* x \gtrsim n_* x^2$ and $n_*^{2-\delta} x^2 \gtrsim n_* x^2$, which implies that

$$\mathbb{P}\{|\mathrm{III}(i,j)| > x\} \leq C \exp(-Cn_* x^2) + Cn_* \exp(-Cn_*^{\vartheta\gamma}) + Cn_* \exp(-Cn_*^{-2\vartheta + \kappa - 1 + \delta})$$
$$\leq C \exp(-Cn_* x^2) + C \exp(-Cn_*^{\vartheta\gamma}) + C \exp(-Cn_*^{-2\vartheta + \kappa - 1 + \delta})$$

for any $0 < x \leq C$. Define $\aleph_2 = \min\{(\vartheta\gamma - 1)/2, (-2\vartheta + \kappa - 2 + \delta)/2\}$. We have

$$\max_{1 \leq i,j \leq p} \mathbb{P}\{|\mathrm{III}(i,j)| > x\} \leq C \exp(-Cn_* x^2)$$

for any $x = o(n_*^{\aleph_2})$. To make $\aleph_2$ be largest, we select $\delta = 1$ and $\vartheta = \kappa/(\gamma + 2)$. Then the corresponding $\aleph_2 = (\kappa\gamma - \gamma - 2)/(2\gamma + 4)$. Similarly, we can show the same result holds for $\mathrm{IV}(i,j)$. We complete the proof of Lemma 3. □

28

6.2 Proof of Theorem 2

To prove Theorem 2, we need the Le Cam's lemma as stated in Lemma 4 below. Its proof can be found in Le Cam (1973) and Donoho and Liu (1991). Let $\mathcal{Z}$ be an observation from a distribution $\mathbb{P}_\theta$ where θ belongs to a parameter space Θ. For two distributions $\mathbb{Q}_0$ and $\mathbb{Q}_1$ with densities q_0 and q_1 with respect to any common dominating measure μ, the total variation affinity is given by $\|\mathbb{Q}_0 \wedge \mathbb{Q}_1\| = \int q_0 \wedge q_1 \, d\mu$. Let $\Theta = \{\theta_0, \theta_1, \ldots, \theta_D\}$ and denote by L the loss function. Define $l_{\min} = \min_{1 \le d \le D} \inf_t \{L(t, \theta_0) + L(t, \theta_d)\}$ and denote $\bar{\mathbb{P}} = D^{-1} \sum_{d=1}^{D} \mathbb{P}_{\theta_d}$.

Lemma 4. (Le Cam's lemma) *Let T be any estimator of θ based on an observation $\mathcal{Z}$ from a distribution $\mathbb{P}_\theta$ with $\theta \in \Theta = \{\theta_0, \theta_1, \ldots, \theta_D\}$, then*

$$\sup_{\theta \in \Theta} \mathbb{E}_{\mathcal{Z}|\theta} \{L(T, \theta)\} \ge \frac{1}{2} l_{\min} \|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\| \,.$$

For each $k = 1, \ldots, n$, define $\mathcal{C}_k = \{1 \le i \le p : t_k \in \mathcal{G}_i\}$ where $\mathcal{G}_i$ is the grid of time points where we observe the noisy data of the ith component process. For any s-dimensional vector $\mathbf{a}$ and an index set $\mathcal{C} \subset \{1, \ldots, s\}$, denote by $\mathbf{a}_\mathcal{C}$ the subvector of $\mathbf{a}$ with components indexed by $\mathcal{C}$. The data we have is $\mathcal{Z} = \{\mathbf{Y}_{t_1, \mathcal{C}_1}, \ldots, \mathbf{Y}_{t_n, \mathcal{C}_n}\}$. Select the loss function $L(T, \theta) = \max_{1 \le i, j \le p} |\omega_{i,j} - \theta_{i,j}|$ for any $T = (\omega_{i,j})_{p \times p}$ and $\theta = (\theta_{i,j})_{p \times p} \in \Theta$. Select $D = p$, $\theta_0 = \mathbf{\Sigma}_{\mathbf{u}, 0} = \mathbf{I}_p$ and

$$\theta_d = \mathbf{\Sigma}_{\mathbf{u}, d} = \mathbf{I}_p + (v^{1/2} n_*^{-1/2} \log^{1/2} D) \operatorname{diag}(\underbrace{0, \ldots, 0}_{d-1}, 1, \underbrace{0, \ldots, 0}_{p-d}) \,,$$

for any $d = 1, \ldots, D$, where $v > 0$ is a sufficiently small constant. For each $d = 0, 1, \ldots, D$, we write $\theta_d = \mathbf{\Sigma}_{\mathbf{u}, d} = (\sigma_{\mathbf{u}, i, j, d})_{p \times p}$. Then

$$
\begin{aligned}
l_{\min} &= \min_{1 \le d \le D} \inf_t \{L(t, \theta_0) + L(t, \theta_d)\} \\
&\ge \min_{1 \le d \le D} \max_{1 \le i, j \le p} |\sigma_{\mathbf{u}, i, j, 0} - \sigma_{\mathbf{u}, i, j, d}| \ge \sqrt{\frac{v \log p}{n_*}} \asymp \sqrt{\frac{\log p}{n_*}} \,.
\end{aligned}
\tag{6.15}
$$

To prove the lower bound stated in Theorem 2, it suffices to construct a specific model which makes the stated lower bound be achievable. To do this, we select $\mu_{i,t} = 0$ and $\sigma_{i,t} = 0$ for any $t \in [0, T]$. Then the associated $\mathbf{X}_t = \mathbf{0}$ for any $t \in [0, T]$. In this special case, $\mathbf{Y}_{t_k} = \mathbf{U}_{t_k}$. Given (n, n_*) with $n \ge n_*$, and $0 \le t_1 < \cdots < t_n = T$, we define $\mathcal{G}_* = \{\tilde{t}_1, \ldots, \tilde{t}_{n_*}\}$ with each $\tilde{t}_j \in \{t_1, \ldots, t_n\}$ and $\tilde{t}_j < \tilde{t}_{j+1}$. For each $t_j \in \mathcal{G}_*$, we assume all p component processes are observed. For any $t_j \notin \mathcal{G}_*$, we assume only one component process are observed. Without lose of generality, we assume $\mathcal{G}_* = \{t_1, \ldots, t_{n_*}\}$. Let $n - n_* = ap + q$ where $a \ge 0$ and $0 \le q < p$ are two integers. We assume the ith component process is observed at t_{n_*+jp+i}'s with $j = 0, \ldots, a$ and $i = 1, \ldots, p$.

Then $\mathcal{G}_i = \mathcal{G}_* \cup \{t_{n_*+i}, \ldots, t_{n_*+ap+i}\}$. It follows from Assumption 1 that $\mathcal{Z} = \{\mathbf{Y}_{t_1,\mathcal{C}_1}, \ldots, \mathbf{Y}_{t_n,\mathcal{C}_n}\}$ are independent observations. Let $\mathbf{U}_{t_k} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{u},d})$, and denote the joint density of $\mathbf{U}_{t_1,\mathcal{C}_1}, \ldots, \mathbf{U}_{t_n,\mathcal{C}_n}$ by f_d. Denote by ϕ_σ the density of $N(0, \sigma^2)$. Write $\sigma_*^2 = 1 + v^{1/2} n_*^{-1/2} \log^{1/2} D$. Then

$$f_0 = \prod_{k=1}^{n} \prod_{j \in \mathcal{C}_k} \phi_1(u_{k,j})$$

and

$$f_d = \prod_{k=1}^{n} \prod_{j \in \mathcal{C}_k \setminus \{d\}} \phi_1(u_{k,j}) \cdot \prod_{k=1}^{n} \prod_{d \in \mathcal{C}_k} \phi_{\sigma_*}(u_{k,d})$$

for each $d = 1, \ldots, D$. Here we adopt the convention $\prod_{d \in \mathcal{C}_k} \phi_{\sigma_*}(u_{k,d}) \equiv 1$ if $d \notin \mathcal{C}_k$. We will show $\|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\| \geq c$ for some uniform constant $c > 0$.

For any two densities q_0 and q_1, by Cauchy-Schwarz inequality, we have

$$\left(\int |q_0 - q_1| \, \mathrm{d}\mu \right)^2 = \left(\int \frac{|q_0 - q_1|}{q_1^{1/2}} q_1^{1/2} \, \mathrm{d}\mu \right)^2 \leq \int \frac{(q_0 - q_1)^2}{q_1} \, \mathrm{d}\mu = \int \frac{q_0^2}{q_1} \, \mathrm{d}\mu - 1 \, ,$$

which implies that

$$\int q_0 \wedge q_1 \, \mathrm{d}\mu = 1 - \frac{1}{2} \int |q_0 - q_1| \, \mathrm{d}\mu \geq 1 - \frac{1}{2} \left(\int \frac{q_0^2}{q_1} \, \mathrm{d}\mu - 1 \right)^{1/2} .$$

To show $\|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\| \geq c$, it suffices to show that $\int (D^{-1} \sum_{d=1}^{D} f_d)^2 f_0^{-1} \, \mathrm{d}\mu - 1 \to 0$, that is,

$$\frac{1}{D^2} \sum_{d=1}^{D} \left(\int \frac{f_d^2}{f_0} \, \mathrm{d}\mu - 1 \right) + \frac{1}{D^2} \sum_{d_1 \neq d_2} \left(\int \frac{f_{d_1} f_{d_2}}{f_0} \, \mathrm{d}\mu - 1 \right) \to 0 \, . \tag{6.16}$$

Notice that

$$\frac{f_{d_1} f_{d_2}}{f_0} = \prod_{k=1}^{n} \left\{ \prod_{d_1 \in \mathcal{C}_k} \phi_{\sigma_*}(u_{k,d_1}) \right\} \left\{ \prod_{d_2 \in \mathcal{C}_k} \phi_{\sigma_*}(u_{k,d_2}) \right\} \left[\prod_{j \in \mathcal{C}_k \setminus \{d_1, d_2\}} \phi_1(u_{k,j}) \right]$$

for any $d_1 \neq d_2$, then $\int f_{d_1} f_{d_2} / f_0 \, \mathrm{d}\mu = 1$, which implies

$$\frac{1}{D^2} \sum_{d_1 \neq d_2} \left(\int \frac{f_{d_1} f_{d_2}}{f_0} \, \mathrm{d}\mu - 1 \right) = 0 \, .$$

For any $d = 1, \ldots, D$, we have

$$\frac{f_d^2}{f_0} = \prod_{k=1}^{n} \prod_{j \in \mathcal{C}_k \setminus \{d\}} \phi_1(u_{k,j}) \cdot \prod_{k=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma_*^2} \exp \left\{ -\frac{(2 - \sigma_*^2) u_{k,d}^2}{2\sigma_*^2} \right\} \right]^{\mathbb{I}(d \in \mathcal{C}_k)} ,$$

30

which implies

$$\int \frac{f_d^2}{f_0}\,\mathrm{d}\mu = \left(\frac{1}{\sigma_*\sqrt{2-\sigma_*^2}}\right)^{\sum_{k=1}^{n}\mathbb{I}(d\in\mathcal{C}_k)} \prod_{k=1}^{n}\prod_{j\in\mathcal{C}_k\setminus\{d\}}\left\{\int \phi_1(u_{k,j})\,\mathrm{d}u_{k,j}\right\}$$

$$\times \prod_{k=1}^{n}\left[\int \frac{\sqrt{2-\sigma_*^2}}{\sqrt{2\pi}\sigma_*}\exp\left\{-\frac{(2-\sigma_*^2)u_{k,d}^2}{2\sigma_*^2}\right\}\mathrm{d}u_{k,d}\right]^{\mathbb{I}(d\in\mathcal{C}_k)}$$

$$= \left(\frac{1}{\sigma_*\sqrt{2-\sigma_*^2}}\right)^{\sum_{k=1}^{n}\mathbb{I}(d\in\mathcal{C}_k)} = \left(1-\frac{v\log D}{n_*}\right)^{-\sum_{k=1}^{n}\mathbb{I}(d\in\mathcal{C}_k)/2}.$$

Notice that $\sum_{k=1}^{n}\mathbb{I}(d\in\mathcal{C}_k)\le n_*+a+1$ for each $d=1,\ldots,D$. Therefore,

$$\int \frac{f_d^2}{f_0}\,\mathrm{d}\mu \le \left(1-\frac{v\log D}{n_*}\right)^{-(n_*+a+1)/2}$$

for each $d = 1,\ldots,D$. Due to $n/n_* = O(p)$, we know $a = O(n_*)$. Applying the inequality $\log(1-x)\ge -2x$ for any $0<x<1/2$, we have

$$0\le \frac{1}{D^2}\sum_{d=1}^{D}\left(\int \frac{f_d^2}{f_0}\,\mathrm{d}\mu - 1\right)\le \exp\left[-\left\{1-v\left(1+\frac{a+1}{n_*}\right)\right\}\log D\right]\to 0$$

for sufficiently small $v>0$. Then (6.16) holds. Hence $\|\mathbb{P}_{\theta_0}\wedge\bar{\mathbb{P}}\|\ge c$ for some uniform constant $c>0$. Together with (6.15), we can obtain Theorem 2 by Lemma 4. $\qquad\square$

6.3 Proof of Theorem 3

For each $i,j=1,\ldots,p$, we define event

$$A_{i,j}=\left[|\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi)-\sigma_{\mathbf{u},i,j}|\le 4\min\left\{|\sigma_{\mathbf{u},i,j}|,\alpha\sqrt{\frac{\log p}{n_*}}\right\}\right]$$

for some constant $\alpha>0$, and

$$d_{i,j}=\{\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi)-\sigma_{\mathbf{u},i,j}\}\mathbb{I}(A_{i,j}^c).$$

Write $\mathbf{D}=(d_{i,j})_{p\times p}$. Due to $\|\mathbf{W}\|_2\le\|\mathbf{W}\|_\infty$ for any $p\times p$ symmetric matrix $\mathbf{W}$, it holds that

$$\|\hat{\mathbf{\Sigma}}_{\mathbf{u}}^{\mathrm{thre}}(\xi)-\mathbf{\Sigma}_{\mathbf{u}}\|_2^2 \le \left\{\max_{1\le i\le p}\sum_{j=1}^{p}|\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi)-\sigma_{\mathbf{u},i,j}|\right\}^2$$

$$\le 2\left(\max_{1\le i\le p}\sum_{j=1}^{p}|d_{i,j}|\right)^2 + 2\left\{\max_{1\le i\le p}\sum_{j=1}^{p}|\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi)-\sigma_{\mathbf{u},i,j}|\mathbb{I}(A_{i,j})\right\}^2. \tag{6.17}$$

In the sequel, we will first bound the second term on the right-hand side of above inequality. Notice that

$$\sum_{j=1}^{p} |\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) - \sigma_{\mathbf{u},i,j}|\mathbb{I}(A_{i,j}) \leq 4\sum_{j=1}^{p} \alpha\sqrt{\frac{\log p}{n_*}}\mathbb{I}\left(|\sigma_{\mathbf{u},i,j}| > \alpha\sqrt{\frac{\log p}{n_*}}\right)$$
$$+ 4\sum_{j=1}^{p} |\sigma_{\mathbf{u},i,j}|\mathbb{I}\left(|\sigma_{\mathbf{u},i,j}| \leq \alpha\sqrt{\frac{\log p}{n_*}}\right).$$

Since $\boldsymbol{\Sigma}_{\mathbf{u}} \in \mathcal{H}(q, c_p, M)$, we know $\sum_{j=1}^{p} |\sigma_{\mathbf{u},i,j}|^q \leq c_p$ for each $i = 1, \ldots, p$, which implies that

$$\sum_{j=1}^{p} |\sigma_{\mathbf{u},i,j}|\mathbb{I}\left(|\sigma_{\mathbf{u},i,j}| \leq \alpha\sqrt{\frac{\log p}{n_*}}\right) \leq \sum_{j=1}^{p} |\sigma_{\mathbf{u},i,j}|^q\left(\alpha\sqrt{\frac{\log p}{n_*}}\right)^{1-q} \leq \alpha^{1-q}c_p\left(\frac{\log p}{n_*}\right)^{(1-q)/2}$$

and

$$\sum_{j=1}^{p} \alpha\sqrt{\frac{\log p}{n_*}}\mathbb{I}\left(|\sigma_{\mathbf{u},i,j}| > \alpha\sqrt{\frac{\log p}{n_*}}\right) \leq \sum_{j=1}^{p} |\sigma_{\mathbf{u},i,j}|^q\left(\alpha\sqrt{\frac{\log p}{n_*}}\right)^{1-q} \leq \alpha^{1-q}c_p\left(\frac{\log p}{n_*}\right)^{(1-q)/2}.$$

Therefore, it holds that

$$\max_{1\leq i\leq p}\sum_{j=1}^{p} |\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) - \sigma_{\mathbf{u},i,j}|\mathbb{I}(A_{i,j}) \leq 8\alpha^{1-q}c_p\left(\frac{\log p}{n_*}\right)^{(1-q)/2}.$$

It follows from (6.17) that

$$\mathbb{E}\left\{\|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\mathrm{thre}}(\xi) - \boldsymbol{\Sigma}_{\mathbf{u}}\|_2^2\right\} \leq 2\mathbb{E}\left\{\left(\max_{1\leq i\leq p}\sum_{j=1}^{p} |d_{i,j}|\right)^2\right\} + 128c_p^2\left(\frac{\log p}{n_*}\right)^{1-q}. \tag{6.18}$$

Recall $\hat{\sigma}_{\mathbf{u},i,j}(\xi)$ is defined as (2.5). On the other hand, we know

$$\mathbb{E}\left\{\left(\max_{1\leq i\leq p}\sum_{j=1}^{p} |d_{i,j}|\right)^2\right\} \leq p\sum_{i,j=1}^{p} \mathbb{E}\left\{|\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) - \sigma_{\mathbf{u},i,j}|^2\mathbb{I}(A_{i,j}^c)\right\}$$
$$= p\underbrace{\sum_{i,j=1}^{p} \mathbb{E}\left(|\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) - \sigma_{\mathbf{u},i,j}|^2\mathbb{I}[A_{i,j}^c \cap \{\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) = 0\}]\right)}_{\mathrm{I}}$$
$$+ p\underbrace{\sum_{i,j=1}^{p} \mathbb{E}\left(|\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) - \sigma_{\mathbf{u},i,j}|^2\mathbb{I}[A_{i,j}^c \cap \{\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) = \hat{\sigma}_{\mathbf{u},i,j}(\xi)\}]\right)}_{\mathrm{II}}.$$

Recall $\hat{\sigma}_{\mathbf{u},i,j}^{\mathrm{thre}}(\xi) = \hat{\sigma}_{\mathbf{u},i,j}(\xi)\mathbb{I}\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)| \geq \beta(n_*^{-1}\log p)^{1/2}\}$ for any $i, j = 1, \ldots, p$. Then

$$\mathrm{I} = p\sum_{i,j=1}^{p} \sigma_{\mathbf{u},i,j}^2\mathbb{P}\left[\left\{|\sigma_{\mathbf{u},i,j}| \geq 4\alpha\sqrt{\frac{\log p}{n_*}}\right\} \cap \left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)| < \beta\sqrt{\frac{\log p}{n_*}}\right\}\right]$$

32

$$\leq p\sum_{i,j=1}^{p}\sigma_{\mathbf{u},i,j}^{2}\mathbb{P}\left[\left\{|\sigma_{\mathbf{u},i,j}|\geq 4\alpha\sqrt{\frac{\log p}{n_{*}}}\right\}\cap\left\{|\sigma_{\mathbf{u},i,j}|-|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|<\beta\sqrt{\frac{\log p}{n_{*}}}\right\}\right]$$

$$\leq p\sum_{i,j=1}^{p}\sigma_{\mathbf{u},i,j}^{2}\mathbb{P}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|\geq(4\alpha-\beta)\sqrt{\frac{\log p}{n_{*}}}\right\}.$$

With $\alpha=\beta/2$ and β being sufficiently large, it follows from (6.1) that $\mathrm{I}\lesssim p^{3-C\beta^{2}}\lesssim n_{*}^{-1}$ provided that $\log p=o\{n_{*}^{\tau(\kappa,\tau)}\}$ with $\tau(\kappa,\tau)=\min\{\kappa/5,(2\kappa-1)\gamma/(\gamma+4)\}$. Also, by Cauchy-Schwarz inequality, it holds that

$$\mathrm{II}=p\sum_{i,j=1}^{p}\mathbb{E}\left(|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|^{2}\mathbb{I}\left[|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|>4\min\left\{|\sigma_{\mathbf{u},i,j}|,\alpha\sqrt{\frac{\log p}{n_{*}}}\right\}\right]\right.$$

$$\left.\times\mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)|\geq\beta\sqrt{\frac{\log p}{n_{*}}}\right\}\right)$$

$$=p\sum_{i,j=1}^{p}\mathbb{E}\left(|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|^{2}\mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|>4\alpha\sqrt{\frac{\log p}{n_{*}}}\right\}\right.$$

$$\left.\times\mathbb{I}\left(|\sigma_{\mathbf{u},i,j}|>\alpha\sqrt{\frac{\log p}{n_{*}}}\right)\mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)|\geq\beta\sqrt{\frac{\log p}{n_{*}}}\right\}\right)$$

$$+p\sum_{i,j=1}^{p}\mathbb{E}\left(|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|^{2}\mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|>4|\sigma_{\mathbf{u},i,j}|\right\}\right.$$

$$\left.\times\mathbb{I}\left(|\sigma_{\mathbf{u},i,j}|\leq\alpha\sqrt{\frac{\log p}{n_{*}}}\right)\mathbb{I}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)|\geq\beta\sqrt{\frac{\log p}{n_{*}}}\right\}\right)$$

$$\leq p\sum_{i,j=1}^{p}\left[\mathbb{E}\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|^{4}\}\right]^{1/2}\left[\mathbb{P}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|>4\alpha\sqrt{\frac{\log p}{n_{*}}}\right\}\right]^{1/2}$$

$$+p\sum_{i,j=1}^{p}\left[\mathbb{E}\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|^{4}\}\right]^{1/2}\left[\mathbb{P}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|>(\beta-\alpha)\sqrt{\frac{\log p}{n_{*}}}\right\}\right]^{1/2}.$$

Notice that $\alpha=\beta/2$. It follows from (6.1) that

$$\max_{1\leq i,j\leq p}\mathbb{P}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|>4\alpha\sqrt{\frac{\log p}{n_{*}}}\right\}$$

$$\leq\max_{1\leq i,j\leq p}\mathbb{P}\left\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|>(\beta-\alpha)\sqrt{\frac{\log p}{n_{*}}}\right\}\lesssim p^{-C\beta^{2}}$$

provided that $\log p=o\{n_{*}^{\tau(\kappa,\tau)}\}$ with $\tau(\kappa,\tau)=\min\{\kappa/5,(2\kappa-1)\gamma/(\gamma+4)\}$, which implies that

$$\mathrm{II}\lesssim p^{3-C\beta^{2}}\max_{1\leq i,j\leq p}\left[\mathbb{E}\{|\hat{\sigma}_{\mathbf{u},i,j}(\xi)-\sigma_{\mathbf{u},i,j}|^{4}\}\right]^{1/2}$$

33

Following the proofs for Lemmas 1–3, we know

$$\max_{1 \le i,j \le p} \big[\mathbb{E}\big\{ |\hat{\sigma}_{\mathbf{u},i,j}(\xi) - \sigma_{\mathbf{u},i,j}|^4 \big\} \big]^{1/2} \le C \,.$$

With sufficiently large β, we have II $\lesssim n_*^{-1}$. Together with I $\lesssim n_*^{-1}$, it holds that

$$\mathbb{E}\bigg\{ \bigg(\max_{1 \le i \le p} \sum_{j=1}^{p} |d_{i,j}| \bigg)^2 \bigg\} \le \mathrm{I} + \mathrm{II} \lesssim n_*^{-1} \,.$$

It follows from (6.18) that

$$\sup_{\mathcal{P}_2} \mathbb{E}\big\{ \|\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{\mathrm{thre}}(\xi) - \boldsymbol{\Sigma}_{\mathbf{u}}\|_2^2 \big\} \lesssim c_p^2 \bigg(\frac{\log p}{n_*} \bigg)^{1-q}$$

provided that $\log p = o\{ n_*^{\tau(\kappa,\tau)} \}$ with $\tau(\kappa,\tau) = \min\{ \kappa/5, (2\kappa - 1)\gamma/(\gamma + 4) \}$. We complete the proof of Theorem 3. $\qquad\square$

6.4   Proof of Theorem 4

Same as the proof of Theorem 2, we also select $\mu_{i,t} = 0$ and $\sigma_{i,t} = 0$ for any $t \in [0, T]$. Then the associated $\mathbf{X}_t = \mathbf{0}$ for any $t \in [0, T]$. In this special case, $\mathbf{Y}_{t_k} = \mathbf{U}_{t_k}$. Given (n, n_*) with $n \ge n_*$, and $0 \le t_1 < \cdots < t_n = T$, we define $\mathcal{G}_* = \{ \tilde{t}_1, \ldots, \tilde{t}_{n_*} \}$ with each $\tilde{t}_j \in \{ t_1, \ldots, t_n \}$ and $\tilde{t}_j < \tilde{t}_{j+1}$. For each $t_j \in \mathcal{G}_*$, we assume all p component processes are observed. For any $t_j \notin \mathcal{G}_*$, we assume only one component process are observed. Without lose of generality, we assume $\mathcal{G}_* = \{ t_1, \ldots, t_{n_*} \}$. Let $n - n_* = ap + q$ where $a \ge 0$ and $0 \le q < p$ are two integers. We assume the ith component process is observed at t_{n_*+jp+i}'s with $j = 0, \ldots, a$ and $i = 1, \ldots, p$. Then $\mathcal{G}_i = \mathcal{G}_* \cup \{ t_{n_*+i}, \ldots, t_{n_*+ap+i} \}$. The data we have is $\mathcal{Z} = \{ \mathbf{Y}_{t_1,\mathcal{C}_1}, \ldots, \mathbf{Y}_{t_n,\mathcal{C}_n} \}$ where $\mathcal{C}_k = \{ 1 \le i \le p : t_k \in \mathcal{G}_i \}$.

Let $r = \lfloor p/2 \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. Let $\mathcal{B}$ be the collection of all p-dimensional row vectors $v = (v_1, \ldots, v_p)$ such that $v_j = 0$ for $1 \le j \le p - r$ and $v_j = 0$ or 1 for $p - r + 1 \le j \le p$ under the restriction $\sum_{j=1}^{p} |v_j| = K$. We will specify K later. If each $\lambda_j \in \mathcal{B}$, we say $\lambda = (\lambda_1, \ldots, \lambda_r) \in \mathcal{B}^r$. Set $\Gamma = \{0, 1\}^r$ and $\Lambda \subset \mathcal{B}^r$. For each $\lambda = (\lambda_1, \ldots, \lambda_r) \in \Lambda$, we define $p \times p$ symmetric matrices $A_1(\lambda_1), \ldots, A_r(\lambda_r)$ where $A_m(\lambda_m)$ is a matrix with the mth row and mth column being λ_m and λ_m^{T}, respectively, and the rest of the entries being 0. Define $\Theta = \Gamma \otimes \Lambda$. For each $\theta \in \Theta$, we write $\theta = \{\gamma(\theta), \lambda(\theta)\}$ with $\gamma(\theta) = \{\gamma_1(\theta), \ldots, \gamma_r(\theta)\} \in \Gamma$ and $\lambda(\theta) = \{\lambda_1(\theta), \ldots, \lambda_r(\theta)\} \in \Lambda$. We select $K = \lfloor c_p(n_*/\log p)^{q/2} \rfloor$ and define a collection $\mathcal{M}(\alpha, \nu)$ of covariance matrices as

$$\mathcal{M}(\alpha, \nu) = \bigg\{ \boldsymbol{\Sigma}(\theta) : \boldsymbol{\Sigma}(\theta) = \alpha \mathbf{I}_p + \sqrt{\frac{\nu \log p}{n_*}} \sum_{m=1}^{r} \gamma_m(\theta) A_m\{\lambda_m(\theta)\}, \ \theta \in \Theta \bigg\},$$

34

where $\alpha > 0$ and $\nu > 0$ are two constants. Notice that each $\mathbf{\Sigma} \in \mathcal{M}(\alpha, \nu)$ has value α along the main diagonal, and contains an $r \times r$ submatrix, say A, at the upper right corner, A^{T} at the lower left corner and 0 elsewhere. Write $\mathbf{\Sigma}(\theta) = \{\sigma_{i,j}(\theta)\}_{p \times p}$. It holds that

$$\max_{\theta \in \Theta} \max_{1 \leq i \leq p} \sigma_{i,i}(\theta) = \alpha \quad \text{and} \quad \max_{\theta \in \Theta} \max_{1 \leq i \leq p} \sum_{j=1}^{p} |\sigma_{i,j}(\theta)|^q \leq \alpha^q + c_p \nu^{q/2}.$$

For sufficiently small α and ν, we have $\mathcal{M}(\alpha, \nu) \subset \mathcal{H}(q, c_p, M)$ for $\mathcal{H}(q, c_p, M)$ defined as (2.7). Without lose of generality, we assume $\alpha = 1$ in the sequel and write $\mathcal{M}(1, \nu)$ as $\mathcal{M}$ for simplification.

Let $\mathbf{U}_{t_k} \sim N\{\mathbf{0}, \mathbf{\Sigma}(\theta)\}$ with $\mathbf{\Sigma}(\theta) \in \mathcal{M}$. When $\mathbf{U}_{t_k} \sim N\{\mathbf{0}, \mathbf{\Sigma}(\theta)\}$, we write the distribution of $\mathcal{Z}$ as $\mathbb{P}_\theta$. More specifically, the joint density of $\mathcal{Z}$ is

$$f_\theta = \prod_{k=1}^{n_*} (2\pi)^{-p/2} [\det\{\mathbf{\Sigma}(\theta)\}]^{-1/2} \exp\left\{ -\frac{1}{2} \mathbf{u}_k^{\mathrm{T}} \mathbf{\Sigma}^{-1}(\theta) \mathbf{u}_k \right\}$$

$$\times \prod_{k=n_*+1}^{n} \{2\pi\sigma_{k,k}(\theta)\}^{-1/2} \exp\left\{ -\frac{u_{k,k}^2}{2\sigma_{k,k}(\theta)} \right\}$$

$$= \prod_{k=1}^{n_*} (2\pi)^{-p/2} [\det\{\mathbf{\Sigma}(\theta)\}]^{-1/2} \exp\left\{ -\frac{1}{2} \mathbf{u}_k^{\mathrm{T}} \mathbf{\Sigma}^{-1}(\theta) \mathbf{u}_k \right\}$$

$$\times \prod_{k=n_*+1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_{k,k}^2}{2} \right)$$

where $\mathbf{u}_k = (u_{k,1}, \ldots, u_{k,p})^{\mathrm{T}}$. It follows from Lemma 3 of Cai and Zhou (2012b) with $s = 2$ and d being the matrix spectral norm $\|\cdot\|_2$ that

$$\inf_{\widehat{\mathbf{\Sigma}}} \max_{\theta \in \Theta} \mathbb{E}_{\mathcal{Z}|\theta}\left\{ \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}(\theta)\|_2^2 \right\}$$

$$\geq \min_{\{(\theta,\theta'):H\{\gamma(\theta),\gamma(\theta')\}\geq 1\}} \frac{\|\mathbf{\Sigma}(\theta) - \mathbf{\Sigma}(\theta')\|_2^2}{H\{\gamma(\theta), \gamma(\theta')\}} \cdot \frac{r}{8} \cdot \min_{1 \leq i \leq r} \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\|,$$

where $H(\cdot, \cdot)$ is the Hamming distance, and

$$\bar{\mathbb{P}}_{i,a} = \frac{1}{2^{r-1}|\Lambda|} \sum_{\theta \in \{\theta \in \Theta : \gamma_i(\theta) = a\}} \mathbb{P}_\theta$$

for each $a \in \{0, 1\}$. In the sequel, we will show the following two results:

$$\min_{\{(\theta,\theta'):H\{\gamma(\theta),\gamma(\theta')\}\geq 1\}} \frac{\|\mathbf{\Sigma}(\theta) - \mathbf{\Sigma}(\theta')\|_2^2}{H\{\gamma(\theta), \gamma(\theta')\}} \gtrsim \frac{c_p^2}{p} \left(\frac{\log p}{n_*} \right)^{1-q} \tag{6.19}$$

and

$$\min_{1 \leq i \leq r} \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\| \gtrsim 1. \tag{6.20}$$

Recall $r = \lfloor p/2 \rfloor$. Then we will have Theorem 4. The proofs of (6.19) and (6.20) are identical to that for Lemmas 5 and 6 in Cai and Zhou (2012b), respectively. Hence, we omit here. □

35

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817–858.

Aït-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.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y. and Yu, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, 3, 422–457.

Bickel, P. J. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36, 199–227.

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

Bolger, N. and Laurenceau (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. The Guilford Press.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106, 672–684.

Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38, 2118–2144.

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

Cai, T. T. and Zhou, H. H. (2012b). Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics*, 40, 2389–2420.

Chang, J., Delaigle, A., Hall, P. and Tang, C. Y. (2018). A frequency domain analysis of the error distribution from noisy high-frequency data. *Biometrika*, 105, 353–369.

Chang, J., Tang, C. Y. and Wu, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *The Annals of Statistics*, 41, 2132–2148.

Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence, II. *The Annals of Statistics*, 19, 633–667.

Dufrenot, G., Jawadi, F. and Louhichi, W. (2014). *Market Microstructure and Nonlinear Dynamics: Keeping Financial Crisis in Context*. Springer, Berlin.

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 & Economics Statistics*, 34, 489–503.

Fan, J., Li, Y. and Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, 107, 412–428.

Jacod, J., Li, Y. and Zheng, X. (2017). Statistical properties of microstructure noise. *Econometrica*, 85, 1133–1174.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *The Annals of Statistics*, 37, 4254–4278.

Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, 1, 38–53.

Li, Y., Xie, S. and Zheng, X. (2016). Efficient estimation of integrated volatility incorporating trading information. *Journal of Econometrics*, 195, 33–50.

Liu, C. and Tang, C. Y. (2014). A quasi-maximum likelihood approach for integrated covariance matrix estimation with high frequency data. *Journal of Econometrics*, 180, 217–232.

Petrov, V. V. (1995). *Limit theorems of probability theory: sequences of independent random variables*. New York: Oxford University Press.

Rothman, A. J., Levina, E. and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104, 177–186.

Saulis, L. and Statulevičius, V. A. (1991). *Limit Theorems for Large Deviations*. Springer Netherlands.

Tao, M., Wang, Y. and Zhou, H. (2013). Optimal sparse volatility matrix estimation for high dimensional itô processes with measurement errors. *The Annals of Statistics*, 41, 1816–1864.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3, 257–295.

Xia, N. and Zheng, X. (2018). On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations. *The Annals of Statistics*, 46, 500–525.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159, 235–250.

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.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, 44, 2281–2321.