



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Optimal covariance matrix estimation for high-dimensional noise in high-frequency data



Jinyuan Chang ^{a,b,*}, Qiao Hu ^b, Cheng Liu ^c, Cheng Yong Tang ^d

^a Guizhou Key Laboratory of Big Data Statistical Analysis, Guizhou University of Finance and Economics, Guiyang, China

^b Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, China

^c School of Economics and Management, Wuhan University, Wuhan, China

^d Department of Statistical Science, Temple University, Philadelphia, PA, USA

ARTICLE INFO

Article history:

Received 30 May 2021

Received in revised form 16 May 2022

Accepted 10 June 2022

Available online 16 September 2022

JEL classification:

C13

C55

C58

Keywords:

High-dimensional covariance matrix

High-frequency data analysis

Measurement error

Minimax optimality

Thresholding

ABSTRACT

We consider high-dimensional measurement errors with high-frequency data. Our objective is on recovering the high-dimensional cross-sectional covariance matrix of the random errors with optimality. In this problem, not all components of the random vector are observed at the same time and the measurement errors are latent variables, leading to major challenges besides high data dimensionality. We propose a new covariance matrix estimator in this context with appropriate localization and thresholding, and then conduct a series of comprehensive theoretical investigations of the proposed estimator. By developing a new technical device integrating the high-frequency data feature with the conventional notion of α -mixing, our analysis successfully accommodates the challenging serial dependence in the measurement errors. Our theoretical analysis establishes the minimax optimal convergence rates associated with two commonly used loss functions; and we demonstrate with concrete cases when the proposed localized estimator with thresholding achieves the minimax optimal convergence rates. Considering that the variances and covariances can be small in reality, we conduct a second-order theoretical analysis that further disentangles the dominating bias in the estimator. A bias-corrected estimator is then proposed to ensure its practical finite sample performance. We also extensively analyze our estimator in the setting with jumps, and show that its performance is reasonably robust. We illustrate the promising empirical performance of the proposed estimator with extensive simulation studies and a real data analysis.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

High-frequency data broadly refer to those collected at time points with very small time intervals between consecutive observations. Exemplary scenarios with high-frequency data include longitudinal observations with intensive repeated measurements (Bolger and Laurenceau, 2013), the tick-by-tick trading data in finance (Zhang et al., 2005), and functional data with dense observations (Zhang and Wang, 2016). High-frequency data are commonly contaminated by some noise, broadly termed as the measurement errors. For measurement errors in the context of functional data analysis, we refer

* Corresponding author at: Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, China.

E-mail addresses: changjinyuan@swufe.edu.cn (J. Chang), huqiao@smail.swufe.edu.cn (Q. Hu), chengliu_eco@whu.edu.cn (C. Liu), yongtang@temple.edu (C.Y. Tang).

to the review article Wang et al. (2016) and reference therein. In high-frequency financial data, as another example, the microstructure noise is well known; see the monograph Ait-Sahalia and Jacod (2014) for an overview.

Despite the central interests on recovering the signals contaminated by the noise, the properties of the noise themselves are of their own great interests. Recently, Jacod et al. (2017) highlighted the importance of statistical properties of the microstructure noise and studied the estimation of its moments; see the recent study of Li and Linton (2022) on the limiting distributions in a broad setting. Chang et al. (2018) investigated recovering the distribution of the noise with some frequency-domain analysis. In a simultaneous and independent work of ours, Da and Xiu (2021) investigated the auto-covariance of the measurement errors with a semiparametric approach that utilizing a working moving-average model. These aforementioned studies aimed on univariate cases. Ubukata and Oya (2009) considered the covariance estimation and testing for measurement errors in a bivariate case. Christensen et al. (2013) proposed an estimator for the covariance matrix of the noise vector in high-frequency finance data. Both Ubukata and Oya (2009) and Christensen et al. (2013) handled fixed dimensional cases with m -dependent or independent measurement errors.

We are motivated to concentrate on high-dimensional cases in this study that shed light on influential practical applications where covariances between different components of the noise could bring us useful information in solving various problems. For example, for functional-type observations, the covariations between the measurement errors may help identifying the common source or reasons of contaminations so that improvement can be developed in designing future investigations. For financial data, such covariations in the high-dimensional microstructure noise may help in better understanding the trading behaviors that may show substantial different pattern between equities. Indeed, Li et al. (2016) found that a parametric function incorporating the market information may account for a substantial contribution to the variations in the microstructure noise. Nevertheless, studying the covariance between different components of the high-dimensional noise in high-frequency data remains little explored.

Our primary interests in this study are on the validity and optimality of the covariance matrix estimation procedure for the high-dimensional noise in high-frequency data. This problem has unique challenges from multiple aspects. First, since the noise of interest are not directly observable, the targeted random vectors are latent. Second, the latency arises together with high data dimensionality and high sampling frequency, two challenging features that interrelates to each other in this investigation. The high-dimensional noise sequence is expected to contain some serial dependence, posing a major methodological and theoretical challenge. The properties of high-dimensional covariance matrix estimation have not yet been explored in this important scenario. Third, 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 these data features affect the statistical properties on the validity and optimality of the covariance matrix estimation remains unclear.

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 et al. (2009), Cai et al. (2010), Cai and Liu (2011), 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 noise; see, for example, Ait-Sahalia et al. (2010), Fan et al. (2012), Tao et al. (2013), Liu and Tang (2014), Lam et al. (2017), and Xia and Zheng (2018). However, it remains little explored on the high-dimensional covariance matrix of the noises in high-frequency data, accommodating all aforementioned challenging features.

Our study makes several contributions to the area. To our best knowledge, our method is the first handling covariance matrix estimation of the serially dependent high-dimensional noises in high-frequency data. Methodologically, to overcome the difficulties due to the latency, asynchronicity, and serially dependent observations, we propose a new approach with appropriate localization and thresholding. Theoretically, to our best knowledge, our technical device integrating high-frequency serial dependence with the α -mixing is a new development of the current state of knowledge; and it can be more broadly applied for solving this class of problems. Meanwhile, 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 establish cases when 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. 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. Furthermore, our theory includes a second-order analysis revealing the dominating bias of the estimator. We then propose a bias-corrected estimator and show that removing such a bias leads to more promising performance, especially when components in the covariance matrix are small. Our analysis also indicates that the proposed localized estimator is robust to the setting with jumps.

The rest of this paper is organized as follows. The methodology is outlined in Section 2, followed by theoretical development in Section 3. Section 4 presents the theory and method handling the situation when the level of the noises is small. Section 5 investigates the robustness of our method in the setting with jumps. Numerical studies with simulation and a real data analysis are presented in Section 6. Section 7 includes some discussions. All technical proofs are given in Section 8. Some additional numerical results are presented in the supplementary material.

2. Methodology

2.1. Model and data

We introduce some notations first. For any positive integer q , we write $[q] := \{1, \dots, q\}$. For a matrix $\mathbf{B} = (b_{i,j})_{s_1 \times s_2}$, let $\mathbf{B}^{\otimes 2} = \mathbf{B}\mathbf{B}^\top$, $\|\mathbf{B}\|_\infty = \max_{i \in [s_1], j \in [s_2]} |b_{i,j}|$, $\|\mathbf{B}\|_1 = \max_{j \in [s_2]} \sum_{i=1}^{s_1} |b_{i,j}|$, $\|\mathbf{B}\|_\infty = \max_{i \in [s_1]} \sum_{j=1}^{s_2} |b_{i,j}|$ and $\|\mathbf{B}\|_2 = \lambda_{\max}^{1/2}(\mathbf{B}^{\otimes 2})$, where $\lambda_{\max}(\mathbf{B}^{\otimes 2})$ denotes the largest eigenvalue of $\mathbf{B}^{\otimes 2}$. Denote by $I(\cdot)$ 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 exist a positive constant c and a large enough integer n_0 such that $a_n/b_n \leq c$ for all $n \geq n_0$. We write $a_n \asymp b_n$ if and only if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously.

The setting of our study contains the signal part — a p -dimensional continuous-time process $(\mathbf{X}_t)_{t \in [0, T]}$, where, without loss of generality, $[0, T]$ is the time frame in which the high-frequency data are observed. We begin with a setting that $\mathbf{X}_t = (X_{1,t}, \dots, X_{p,t})^\top$ satisfies:

$$dX_{i,t} = \mu_{i,t} dt + \sigma_{i,t} dB_{i,t} \quad \text{and} \quad \mathbb{E}(dB_{i,t} \cdot dB_{j,t}) = \rho_{i,j,t} dt, \tag{1}$$

where $\mu_{i,t}$ and $\sigma_{i,t}$ are progressively measurable processes, and $B_{1,t}, \dots, B_{p,t}$ are univariate standard Brownian motions. Here $\sigma_{i,t}$ and $\rho_{i,j,t}$ are, respectively, governing the volatilities and correlations, where both of them may be dynamic over time. A theoretical study of our method in the setting with jumps will be considered in Section 5.

For each $i \in [p]$, we use $\mathcal{G}_i = \{t_{i,1}, \dots, t_{i,n_i}\}$ to denote the grid of time points at which we observe the noisy data of the i th component process $X_{i,t}$, where $0 \leq t_{i,1} < \dots < t_{i,n_i} \leq T$. 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 $0 \leq t_1 < \dots < t_n \leq T$. For any $i, j \in [p]$, 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 i th and j th component processes $X_{i,t}$ and $X_{j,t}$ simultaneously. Clearly, $n_{i,i} = n_i$ for any $i \in [p]$.

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

$$Y_{i,t_k} = X_{i,t_k} + U_{i,t_k}$$

with $\mathbb{E}(U_{i,t_k}) = 0$ for each $i \in [p]$ and $k \in [n_i]$. The additive noise assumption is common in the literature; see [Ait-Sahalia and Jacod \(2014\)](#). Formally, we can write

$$\mathbf{Y}_{t_k} = \mathbf{X}_{t_k} + \mathbf{U}_{t_k}, \quad k \in [n], \tag{2}$$

and assume the measurement errors $\{\mathbf{U}_{t_k}\}_{k=1}^n$ are independent of the process $(\mathbf{X}_t)_{t \in [0, T]}$. At each time point t_k , we only observe $\sum_{i=1}^p I(t_k \in \mathcal{G}_i)$ components of \mathbf{Y}_{t_k} .

Besides the cross-sectional dependence, serial dependence is expected to be the case for $\{\mathbf{U}_{t_k}\}_{k=1}^n$; our study accommodates such a feature with an innovative device. Denote by $\mathcal{F}_{-\infty}^s$ and \mathcal{F}_s^∞ the σ -fields generated by $\{\mathbf{U}_{t_k}\}_{k \leq s}$ and $\{\mathbf{U}_{t_k}\}_{k \geq s}$, respectively, the α -mixing coefficients are defined as

$$\alpha_n(m) = \sup_s \sup_{A \in \mathcal{F}_{-\infty}^s, B \in \mathcal{F}_{s+m}^\infty} |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|, \quad m \geq 1. \tag{3}$$

Then $\{\mathbf{U}_{t_k}\}_{k=1}^n$ is an α -mixing sequence if $\alpha_n(m) \rightarrow 0$ as $m \rightarrow \infty$. The notion of α -mixing is a conventional foundation for broadly characterizing the serial dependence. Among others, causal ARMA processes with continuous innovation distributions are α -mixing with exponentially decaying α -mixing coefficients, so are stationary Markov chains satisfying certain conditions; see Section 2.6.1 of [Fan and Yao \(2003\)](#). Stationary GARCH models with finite second moments and continuous innovation distributions are also α -mixing with exponentially decaying α -mixing coefficients; see Proposition 12 of [Carrasco and Chen \(2002\)](#). Under certain conditions, vector auto-regressive (VAR) processes, multivariate ARCH processes, and multivariate GARCH processes are all α -mixing with exponentially decaying α -mixing coefficients; see [Hafner and Preminger \(2009\)](#), [Boussama et al. \(2011\)](#) and [Wong et al. \(2020\)](#).

In (3), we highlight the necessary inclusion of n , the frequency related sample size, in the α -mixing coefficient. The reason is that in a high-dimensional data setting, p is commonly specified as a function of the sample size n . Such an intrinsic dependence makes characterizing the serial dependence substantially more challenging. To handle it in our study, we impose the following assumption on $\alpha_n(m)$ defined in (3).

Assumption 1. There exist some universal constants $C_1 > 1$, $C_2 > 0$ and $\varphi > 0$ such that $\alpha_n(m) \leq C_1 \exp\{-C_2(L_n^{-1}m)^\varphi\}$ for any $m \geq 1$, where $L_n > 0$ may diverge with n .

Assumption 1 is our new dedicated device for characterizing the serial dependence of $\{\mathbf{U}_{t_k}\}_{k=1}^n$ in the context of high-frequency high-dimensional data. Here L_n is introduced as a parameter to handle the aforementioned challenge due to the high data dimensionality, together with the conventional m as in the α -mixing settings for analyzing time series. As a

development of its own interests, the synthetic device in [Assumption 1](#) successfully integrates the considerations of high-frequency and high-dimensional data, where the usual interpretation of the α -mixing remains: the between-observation dependence is still getting weaker when they are further away in the serial data, as characterized by both L_n and m . Intuitively, the rationale is that L_n , as a standalone parameter, may diverge together with the sampling frequency and data dimensionality in a synthetic manner. Such a divergence reflects the nature of this more challenging problem due to relatively limited data information, in the sense that the serial dependence in the measurement errors will become stronger as L_n increases.

More specifically, [Assumption 1](#) does not require $\{\mathbf{U}_{t_k}\}$ to be strictly stationary, and it includes several commonly used models for $\{\mathbf{U}_{t_k}\}$ as special cases. For an independent sequence $\{\mathbf{U}_{t_k}\}$, we can select $L_n = 1/2$ and $\varphi = \infty$ in [Assumption 1](#). For an L_n -dependent sequence $\{\mathbf{U}_{t_k}\}$, we can select $\varphi = \infty$ in [Assumption 1](#). If $\{\mathbf{U}_{t_k}\}$ follows VAR model, multivariate ARCH model or multivariate GARCH model with certain conditions, we can select $L_n = \varphi = 1$ in [Assumption 1](#). We provide a concrete example here with a diverging L_n . For each $i \in [r]$, let $Z_{i,t}$ satisfy the diffusion process $dZ_{i,t} = \tilde{\mu}_i(Z_{i,t}; \boldsymbol{\theta}_i) dt + \tilde{\sigma}_i(Z_{i,t}; \boldsymbol{\gamma}_i) dW_{i,t}$, where $W_{i,t}$ is a univariate standard Brownian motion, $\tilde{\mu}_i(\cdot; \cdot)$ and $\tilde{\sigma}_i(\cdot; \cdot)$ are two functions of $Z_{i,t}$ with some parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\gamma}_i$, respectively. Write $\mathbf{Z}_t = (Z_{1,t}, \dots, Z_{r,t})^\top$ with r independent processes $Z_{1,t}, \dots, Z_{r,t}$. Letting $\mathbf{U}_{t_k} = \mathbf{A}\mathbf{Z}_{k\delta}$ for some known loading matrix $\mathbf{A} \in \mathbb{R}^{p \times r}$ and some $\delta > 0$, we can select $L_n = \delta^{-1}$ and $\varphi = 1$ when $\tilde{\mu}_i(\cdot; \cdot)$ and $\tilde{\sigma}_i(\cdot; \cdot)$ satisfy certain conditions,¹ where L_n will diverge with n if $\delta \rightarrow 0$ as $n \rightarrow \infty$. Here L_n is also allowed to depend directly on p , the dimension of \mathbf{U}_{t_k} . As an example, if each univariate sequence $\{U_{i,t_k}\}_{k=1}^n$ is α -mixing with exponentially decaying α -mixing coefficients, with the independence assumption imposed on the p sequences $\{U_{1,t_k}\}_{k=1}^n, \dots, \{U_{p,t_k}\}_{k=1}^n$, [Theorem 5.1 of Bradley \(2005\)](#) indicates that $\alpha_n(m)$ defined in (3) satisfies $\alpha_n(m) \leq p \exp(-cm)$ for some universal constant $c > 0$, which implies [Assumption 1](#) holds for $\varphi = 1$ and $L_n \asymp \log p$.

To our best knowledge, there is no alternative assumption in the literature that is capable of handling the setting of our study. In existing studies, some serial dependence assumptions have been imposed on the measurement errors, with a primary objective recovering its auto-covariance. When $p = 1$, [Jacod et al. \(2017\)](#) assume $U_{1,t_k} = \gamma_{t_k} \chi_k$ for some nonnegative semimartingale γ_t and a ρ -mixing stationary sequence $\{\chi_k\}_{k \geq 1}$, where $\{\chi_k\}_{k \geq 1}$ is independent of the process γ_t ; see also the setting of [Li and Linton \(2022\)](#) that covers serially dependent, endogenous, and nonstationary noises. If γ_t is the solution of some stochastic differential equations, $\{\gamma_{t_k}\}_{k \geq 1}$ is also ρ -mixing. See, for example, [Lemma 4 of Ait-Sahalia and Mykland \(2004\)](#). Based on the independence between $\{\chi_k\}_{k \geq 1}$ and $\{\gamma_{t_k}\}_{k \geq 1}$, [Theorem 5.2 of Bradley \(2005\)](#) implies the sequence $\{U_{1,t_k}\}_{k \geq 1}$ is also ρ -mixing. Since ρ -mixing implies α -mixing, we know $\{U_{1,t_k}\}_{k \geq 1}$ is also α -mixing. [Varneskov \(2017\)](#) relaxes the ρ -mixing assumption on $\{\chi_k\}_{k \geq 1}$ to the weaker α -mixing condition. In a recent study, [Da and Xiu \(2021\)](#) assume instead a working moving average structure for the measurement errors.

For (2), we assume $\text{Cov}(\mathbf{U}_{t_k}) \equiv \boldsymbol{\Sigma}_u$ for each $k \in [n]$. Our main goal in this study is to estimate $\boldsymbol{\Sigma}_u$, the covariance matrix that contains information on the between-component relationship of the unobserved noise \mathbf{U}_{t_k} . Clearly, \mathbf{U}_{t_k} is a latent vector. To estimate its covariance matrix, eliminating the impact due to the process \mathbf{X}_t is required, which means that now \mathbf{U}_{t_k} performs like ‘signal’ and \mathbf{X}_{t_k} is ‘noise’. Our strategy is to perform a dedicated localization: focusing on observations that are in a specific neighborhood mentioned later. For any $i, j \in [p]$, we write $\mathcal{G}_i \cap \mathcal{G}_j = \{t_{i,j,1}, \dots, t_{i,j,n_{i,j}}\}$ with $t_{i,j,1} < \dots < t_{i,j,n_{i,j}}$. Let $\Delta t_{i,j,k} = t_{i,j,k+1} - t_{i,j,k}$ for any $k \in [n_{i,j} - 1]$. In this paper, we consider the scenario with T being fixed but $\max_{i,j \in [p]} \max_{k \in [n_{i,j}-1]} \Delta t_{i,j,k} \rightarrow 0$ as $n \rightarrow \infty$. Formally, we make the following assumption:

Assumption 2. (i) As $n \rightarrow \infty$, $\min_{i,j \in [p]} \min_{k \in [n_{i,j}-1]} \Delta t_{i,j,k} / \max_{i,j \in [p]} \max_{k \in [n_{i,j}-1]} \Delta t_{i,j,k}$ is uniformly bounded away from zero. (ii) As $n \rightarrow \infty$, we have each $n_{i,j} \rightarrow \infty$, and $\min_{i,j \in [p]} n_{i,j} / \max_{i,j \in [p]} n_{i,j}$ is uniformly bounded away from zero. (iii) $\min_{i,j \in [p]} (t_{i,j,n_{i,j}} - t_{i,j,1}) \asymp T$.

The setting with [Assumption 2](#) is broad and general. The first part 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 [Ait-Sahalia et al. \(2010\)](#) 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

$$\min_{i,j \in [p]} n_{i,j} \asymp \max_{i,j \in [p]} n_{i,j} \asymp n_* , \tag{4}$$

where $n_* \rightarrow \infty$ as $n \rightarrow \infty$. As we will show in [Theorems 1–4](#), the convergence rates for the estimates of the covariance matrix $\boldsymbol{\Sigma}_u = \text{Cov}(\mathbf{U}_{t_k})$ will depend on n_* instead of n . In the special case with synchronous observations, we have $n_{i,j} = n$ for any $i, j \in [p]$ and we can set $n_* = n$. Then all our results also apply to the setting with synchronous data. [Assumption 2](#) is not necessary for our theoretical analysis which is just imposed for simplicity and can be removed at the expenses of lengthier proofs. Our theoretical analysis essentially only requires the assumption that $\min_{i,j \in [p]} n_{i,j} \rightarrow \infty$ and $\max_{i,j \in [p]} \max_{k \in [n_{i,j}-1]} \Delta t_{i,j,k} \rightarrow 0$ as $n \rightarrow \infty$. With such assumption, both $\min_{i,j \in [p]} \min_{k \in [n_{i,j}-1]} \Delta t_{i,j,k} / \max_{i,j \in [p]} \max_{k \in [n_{i,j}-1]} \Delta t_{i,j,k}$ and $\min_{i,j \in [p]} n_{i,j} / \max_{i,j \in [p]} n_{i,j}$ can decay to zero as $n \rightarrow \infty$. We

¹ For each $i \in [r]$, [Lemma 4 of Ait-Sahalia and Mykland \(2004\)](#) indicates that $\{Z_{i,k\delta}\}_{k \geq 1}$ is a ρ -mixing process with ρ -mixing coefficient $\rho_i(m) \leq \exp(-c_i m \delta)$ for any integer $m \geq 0$, where $c_i > 0$ is a constant depending on the properties of $\tilde{\mu}_i(\cdot; \cdot)$ and $\tilde{\sigma}_i(\cdot; \cdot)$ (see [Assumption 1 of Ait-Sahalia and Mykland \(2004\)](#)). [Theorem 5.1 of Bradley \(2005\)](#) implies $\{\mathbf{Z}_{k\delta}\}_{k \geq 1}$ is also a ρ -mixing process with ρ -mixing coefficient $\rho(m) \leq \exp(-c_{\min} m \delta)$ for any integer $m \geq 0$, where $c_{\min} = \min_{i \in [r]} c_i$. Since ρ -mixing implies α -mixing, then $\alpha_n(m)$ defined in (3) satisfies $\alpha_n(m) \leq 4^{-1} \rho(m) \leq 4^{-1} \exp(-c_{\min} m \delta)$ for any integer $m \geq 0$.

will discuss in Section 7 how this assumption affects the convergence rates for the estimates of the covariance matrix Σ_u .

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

Write $\Sigma_u = (\sigma_{u,i,j})_{p \times p}$ and $\mathbf{U}_{t_k} = (U_{1,t_k}, \dots, U_{p,t_k})^\top$. Here the subscript u in $\sigma_{u,i,j}$ indicates that it is a quantity associated with the noise so as to differentiate it from the volatility process $\sigma_{i,t}$ in (1). We know $2\Sigma_u = \text{Cov}(\mathbf{U}_{t_k} - \mathbf{U}_{t_\ell}) + \mathbb{E}(\mathbf{U}_{t_\ell} \mathbf{U}_{t_k}^\top) + \mathbb{E}(\mathbf{U}_{t_k} \mathbf{U}_{t_\ell}^\top)$ for any $\ell \neq k$. By Assumption 1 and Davydov's inequality (Davydov, 1968), $|\mathbb{E}(\mathbf{U}_{t_\ell} \mathbf{U}_{t_k}^\top)|_\infty + |\mathbb{E}(\mathbf{U}_{t_k} \mathbf{U}_{t_\ell}^\top)|_\infty \lesssim \exp(-C_* L_n^{-\varphi} |k - \ell|^\varphi)$ for some constant $C_* > 0$, provided that $\max_{j \in [p]} \mathbb{E}(|U_{j,t_k}|^\gamma)$ and $\max_{j \in [p]} \mathbb{E}(|U_{j,t_\ell}|^\gamma)$ are uniformly bounded away from infinity for some universal constant $\gamma > 2$. 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 component process $X_{i,t}$ is a continuous-time and continuous-path stochastic process. We have considerations from two ends. First, due to $|X_{i,t+h} - X_{i,t}| \rightarrow 0$ almost surely as $h \rightarrow 0$, in a small neighborhood \mathcal{N} of t_k , the difference between the high-frequency observations \mathbf{Y}_{t_k} and \mathbf{Y}_{t_ℓ} , for $t_\ell \in \mathcal{N}$, can be approximately viewed as $\mathbf{U}_{t_k} - \mathbf{U}_{t_\ell}$. Second, to avoid excessive impact from aggregating $\mathbb{E}(\mathbf{U}_{t_\ell} \mathbf{U}_{t_k}^\top) + \mathbb{E}(\mathbf{U}_{t_k} \mathbf{U}_{t_\ell}^\top)$, we cannot choose t_ℓ and t_k too close. Putting these two considerations together, we propose to estimate $\sigma_{u,i,j}$ by

$$\hat{\sigma}_{u,i,j} = \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (Y_{i,t_{i,j,\ell}} - Y_{i,t_{i,j,k}})(Y_{j,t_{i,j,\ell}} - Y_{j,t_{i,j,k}}), \tag{5}$$

where $S_{i,j,k} = \{t_{i,j,\ell} \in \mathcal{G}_i \cap \mathcal{G}_j : K \leq |\ell - k| \leq K + \Delta_K\}$ for some integers $K \geq 1$ and $\Delta_K \geq 0$, and $N_{i,j,k} = |S_{i,j,k}|$. Here the set $S_{i,j,k}$ is designed to meet the aforementioned two considerations – ensuring data in an appropriate range are incorporated for estimating Σ_u .

Our estimator (5) with $S_{i,j,k}$ is generally applicable. For an independent sequence $\{\mathbf{U}_{t_k}\}$, we can select $K = 1$ and then $\mathbb{E}(U_{i,t_{i,j,\ell}} U_{j,t_{i,j,k}}) = \mathbb{E}(U_{i,t_{i,j,k}} U_{j,t_{i,j,\ell}}) = 0$ for any $t_{i,j,\ell} \in S_{i,j,k}$ due to $(L_n, \varphi) = (1/2, \infty)$ in Assumption 1. For an L_n -dependent sequence $\{\mathbf{U}_{t_k}\}$, we can select $K > L_n$ due to $\varphi = \infty$ in Assumption 1 and then $\mathbb{E}(U_{i,t_{i,j,\ell}} U_{j,t_{i,j,k}}) = \mathbb{E}(U_{i,t_{i,j,k}} U_{j,t_{i,j,\ell}}) = 0$ for any $t_{i,j,\ell} \in S_{i,j,k}$. For general case with $\varphi < \infty$, with selecting $K \geq L_n(C_{**} \log n_*)^{1/\varphi}$ for some sufficiently large constant $C_{**} > 0$, $\max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} |\mathbb{E}(U_{i,t_{i,j,\ell}} U_{j,t_{i,j,k}}) + \mathbb{E}(U_{i,t_{i,j,k}} U_{j,t_{i,j,\ell}})| \lesssim n_*^{-C_{**} C_{**}}$, which is negligible in comparison to the bias from approximating $(U_{i,t_{i,j,\ell}} - \hat{U}_{i,t_{i,j,k}}, U_{j,t_{i,j,\ell}} - \hat{U}_{j,t_{i,j,k}})$ by $(Y_{i,t_{i,j,\ell}} - Y_{i,t_{i,j,k}}, Y_{j,t_{i,j,\ell}} - Y_{j,t_{i,j,k}})$. To simplify our presentation, we assume $\Delta_K \geq 0$ is a fixed integer in this paper. Our theoretical results can be parallel extended to the scenario with diverging Δ_K .

For a fixed T , (4) and Assumption 2 imply that

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

Let

$$\widehat{\Sigma}_u = (\hat{\sigma}_{u,i,j})_{p \times p} \tag{6}$$

for $\hat{\sigma}_{u,i,j}$ defined as (5). Theorem 1 in Section 3 shows that the elements of $\widehat{\Sigma}_u$ are uniformly consistent to the corresponding elements of Σ_u with a suitable selection of K , i.e.

$$\mathbb{E}(|\widehat{\Sigma}_u - \Sigma_u|_\infty) \lesssim (Kn_*^{-1} \log p)^{1/2}.$$

Theorem 2 in Section 3 shows that $(n_*^{-1} \log p)^{1/2}$ 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. If $\{\mathbf{U}_{t_k}\}$ is an independent or L_n -dependent sequence with fixed L_n , we can select K as a fixed integer and then the associated convergence rate of $|\widehat{\Sigma}_u - \Sigma_u|_\infty$ is minimax optimal. For general cases with $\varphi < \infty$ and fixed L_n , with selecting $K \asymp (\log n_*)^{1/\varphi + \epsilon}$ for some $\epsilon > 0$, the convergence rate of $|\widehat{\Sigma}_u - \Sigma_u|_\infty$ is nearly optimal with an additional logarithm factor $(\log n_*)^{1/(2\varphi) + \epsilon/2}$. More importantly, as we will discuss below Remark 3 in Section 3, $(n_*^{-1} \log p)^{1/2}$ is also 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 noise, 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 $\widehat{\Sigma}_u$ may not be consistent to Σ_u under the spectral norm $\|\cdot\|_2$ 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):

$$\mathcal{H}(q, c_p, M) = \left\{ \Sigma_u = (\sigma_{u,i,j})_{p \times p} : \sigma_{u,i,i} \leq M \text{ and } \sum_{j=1}^p |\sigma_{u,i,j}|^q \leq c_p \text{ for all } i \right\}, \tag{7}$$

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 Σ_u , i.e., if c_p is smaller, then Σ_u is more sparse. If $q = 0$, we have

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

where c_p evaluates the number of nonzero components in each row of Σ_u .

For $\Sigma_u \in \mathcal{H}(q, c_p, M)$, we propose the following thresholding estimator based on the element-wise estimation $\widehat{\Sigma}_u$ given in (6):

$$\widehat{\Sigma}_u^{\text{thre}} = [\widehat{\sigma}_{u,i,j} I\{|\widehat{\sigma}_{u,i,j}| \geq \beta(Kn_*^{-1} \log p)^{1/2}\}]_{p \times p}, \tag{8}$$

where $\beta > 0$ is a fixed constant for the thresholding level. **Theorem 3** in Section 3 shows that such defined thresholding estimator $\widehat{\Sigma}_u^{\text{thre}}$ is consistent to Σ_u under the spectral norm with suitable selections of K and β , i.e.

$$\mathbb{E}(\|\widehat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim c_p^2(Kn_*^{-1} \log p)^{1-q}. \tag{9}$$

Furthermore, **Theorem 4** in Section 3 indicates 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, which is also the minimax optimal convergence rate in the spectral norm loss if we have observations of the noise directly. If $\{\mathbf{U}_{t_k}\}$ is an independent or L_n -dependent sequence with fixed L_n , we can select K as a fixed integer and then the associated convergence rate of $\|\widehat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2$ is minimax optimal. For general cases with $\varphi < \infty$ and fixed L_n , with selecting $K \asymp (\log n_*)^{1/\varphi + \epsilon}$ for some $\epsilon > 0$, the convergence rate of $\|\widehat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2$ is nearly optimal with an additional logarithm factor $(\log n_*)^{(1/\varphi + \epsilon)(1-q)/2}$.

Remark 1. In finite samples, the thresholding estimator $\widehat{\Sigma}_u^{\text{thre}}$ given in (8) may not be positive definite in general. We can first apply the singular value decomposition to $\widehat{\Sigma}_u^{\text{thre}}$: $\widehat{\Sigma}_u^{\text{thre}} = \widehat{\mathbf{P}}^\top \text{diag}(\widehat{\tau}_1, \dots, \widehat{\tau}_p) \widehat{\mathbf{P}}$, where $\widehat{\tau}_1 \geq \dots \geq \widehat{\tau}_p$ are the eigenvalues of $\widehat{\Sigma}_u^{\text{thre}}$, and $\widehat{\mathbf{P}}$ is an orthogonal matrix. If there are s negative eigenvalues, we can use $\widehat{\Sigma}_u^{\text{thre}} = \widehat{\mathbf{P}}^\top \text{diag}(\widehat{\tau}_1, \dots, \widehat{\tau}_{p-s}, \widehat{\tau}_{p-s+1} + \epsilon, \dots, \widehat{\tau}_p + \epsilon) \widehat{\mathbf{P}}$ as the estimate of Σ_u for some $\epsilon > 0$. Write $\delta_n = c_p(Kn_*^{-1} \log p)^{(1-q)/2}$ and let $\tau_1 \geq \dots \geq \tau_p > 0$ be the eigenvalues of Σ_u . Since $\max_{j \in [p]} |\widehat{\tau}_j - \tau_j| \leq \|\widehat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2 = O_p(\delta_n)$, if τ_p is uniformly bounded away from zero and we select $\epsilon = -\widehat{\tau}_p + (Kn_*^{-1} \log p)^{1/2}$ when $\widehat{\tau}_p < 0$, such defined $\widehat{\Sigma}_u^{\text{thre}}$ is positive definite and also satisfies (9).

3. Theoretical analysis

In this section, we establish the theoretical properties of the proposed estimators. To mimic the high-dimensional scenario, we always assume $p \geq n_*^\kappa$ for some universal constant $\kappa > 0$ in this paper. We also require the following assumptions.

Assumption 3. Write $\mathbf{U}_{t_k} = (U_{1,t_k}, \dots, U_{p,t_k})^\top$. There exist some universal constants $C_3 > 1$ and $C_4 > 0$ such that $\mathbb{P}(|U_{i,t_k}| > u) \leq C_3 \exp(-C_4 u^2)$ for any $i \in [p]$, $k \in [n]$ and $u > 0$.

Assumption 4. There exist some universal constants $C_5 > 0$, $C_6 > 0$ and $C_7 > 0$ such that (i) $\mathbb{E}(\exp\{\theta(\mu_{i,t}^2 - \mathbb{E}(\mu_{i,t}^2))\}) \leq \exp(C_6 \theta^2)$ and $\mathbb{E}(\exp\{\theta(\sigma_{i,t}^2 - \mathbb{E}(\sigma_{i,t}^2))\}) \leq \exp(C_6 \theta^2)$ for any $i \in [p]$, $t \in [0, T]$ and $|\theta| \leq C_5$; (ii) $\mathbb{E}(\mu_{i,t}^2) \leq C_7$ and $\mathbb{E}(\sigma_{i,t}^2) \leq C_7$ for any $i \in [p]$ and $t \in [0, T]$.

Assumption 5. There exist some universal constants $\gamma > 0$, $C_8 > 1$ and $C_9 > 0$ such that $\mathbb{P}(\sup_{0 \leq t \leq T} \sigma_{i,t} > u) \leq C_8 \exp(-C_9 u^\gamma)$ for any $i \in [p]$ 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 that part (i) of **Assumption 4** holds if there exist two positive constants C_* and C_{**} such that $\mathbb{P}\{|\mu_{i,t}^2 - \mathbb{E}(\mu_{i,t}^2)| \geq u\} \leq C_* \exp(-C_{**}u)$ and $\mathbb{P}\{|\sigma_{i,t}^2 - \mathbb{E}(\sigma_{i,t}^2)| \geq u\} \leq C_* \exp(-C_{**}u)$ for any $i \in [p]$, $t \in [0, T]$ 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 away from infinity over $i \in [p]$ and $t \in [0, T]$, 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 **Assumption 3**, $\mathbf{X}_t = (X_{1,t}, \dots, X_{p,t})^\top$ follows model (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 $K > C_{L_n}$ for some constant $C \geq 1$. Under **Assumption 1**, it holds that

$$\sup_{\mathcal{P}_1} \mathbb{E}(\|\widehat{\Sigma}_u - \Sigma_u\|_\infty) \lesssim (Kn_*^{-1} \log p)^{1/2}$$

provided that $\log p = o[\min\{(n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}, (n_*K^{-1})^\chi\}]$ and $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, where n_* is specified in (4) and $\chi = \min\{\gamma/(\gamma + 4), 1/3\}$.

Remark 2. **Theorem 1** gives the convergence rate of $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}|$.

(i) For an independent sequence $\{\mathbf{U}_{t_k}\}$, due to $(L_n, \varphi) = (1/2, \infty)$ and $K \geq 1$, $\mathbb{E}(|\hat{\Sigma}_u - \Sigma_u|_\infty) \lesssim (Kn_*^{-1} \log p)^{1/2}$ provided that $\log p = o\{(n_*K^{-1})^\chi\}$. For fixed K , $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| = O_p\{(n_*^{-1} \log p)^{1/2}\}$.

(ii) For an L_n -dependent sequence $\{\mathbf{U}_{t_k}\}$, due to $\varphi = \infty$ and $K > L_n$, $\mathbb{E}(|\hat{\Sigma}_u - \Sigma_u|_\infty) \lesssim (Kn_*^{-1} \log p)^{1/2}$ provided that $\log p = o\{(n_*K^{-1})^\chi\}$. For fixed L_n , $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| = O_p\{(n_*^{-1} \log p)^{1/2}\}$ with selecting a fixed $K > L_n$. For diverging L_n , $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| = O_p\{(L_n n_*^{-1} \log p)^{1/2}\}$ with selecting $K \asymp L_n$ and $K > L_n$.

(iii) For the general cases with $\varphi < \infty$ and fixed L_n , to make $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, we need to select $K \gg (\log n_*)^{1/\varphi}$. If we select $K \asymp (\log n_*)^{1/\varphi + \epsilon}$ for some $\epsilon > 0$, $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| = O_p\{(n_*^{-1} \log p)^{1/2} (\log n_*)^{1/(2\varphi) + \epsilon/2}\}$.

Furthermore, **Theorem 2** shows that the convergence rate $(n_*^{-1} \log p)^{1/2}$ 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

$$\inf_{\hat{\Sigma} \in \check{\mathcal{F}}} \sup_{\mathcal{P}_1} \mathbb{E}(|\hat{\Sigma} - \Sigma_u|_\infty) \gtrsim (n_*^{-1} \log p)^{1/2},$$

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

Remark 3. (i) If $\{\mathbf{U}_{t_k}\}$ is an independent sequence or L_n -dependent sequence with fixed L_n , **Remarks 2(i)** and **2(ii)** indicate that our proposed estimate $\hat{\Sigma}_u$ is minimax optimal under the maximum element-wise loss.

(ii) For the general cases with $\varphi < \infty$ and fixed L_n , **Remark 2(iii)** indicates our proposed estimate $\hat{\Sigma}_u$ is nearly minimax optimal under the maximum element-wise loss with an additional logarithm factor $(\log n_*)^{1/(2\varphi) + \epsilon/2}$.

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 $i \in [p]$. Let $C_k = \{i \in [p] : t_k \in \mathcal{G}_i\}$ for any $k \in [n]$. In this specific model, the latent process $\mathbf{X}_t = \mathbf{0}$ for any $t \in [0, T]$ and thus the data we observed are $\mathcal{Z} = \{\mathbf{U}_{t_1, C_1}, \dots, \mathbf{U}_{t_n, C_n}\}$. Here \mathbf{U}_{t_k, C_k} denotes the subvector of \mathbf{U}_{t_k} with components indexed by C_k . Hence, $(n_*^{-1} \log p)^{1/2}$ is also the minimax optimal rate in the maximum element-wise loss for the covariance matrix estimations of \mathbf{U}_{t_k} with data $\mathcal{Z} = \{\mathbf{U}_{t_1, C_1}, \dots, \mathbf{U}_{t_n, C_n}\}$, which indicates that the estimator $\hat{\Sigma}_u$ shares some oracle property and the proposed localization actually makes the impact of the latent process \mathbf{X}_t be negligible.

Regarding the loss function under the spectral norm $\|\cdot\|_2$ for the whole covariance matrix estimation, **Theorem 3** establishes the convergence rate of the thresholding estimator $\hat{\Sigma}_u^{\text{thre}}$ defined as (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 **Assumption 3** with the covariance matrix $\Sigma_u \in \mathcal{H}(q, c_p, M)$, $\mathbf{X}_t = (X_{1,t}, \dots, X_{p,t})^\top$ follows model (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 $K > CL_n$ for some constant $C \geq 1$. Under **Assumption 1**, with sufficiently large constant $\beta > 0$ in (8), it holds that

$$\sup_{\mathcal{P}_2} \mathbb{E}(\|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim c_p^2 (Kn_*^{-1} \log p)^{1-q}$$

provided that $\log p = o[\min\{(n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}, (n_*K^{-1})^\chi\}]$ and $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, where n_* is specified in (4) and $\chi = \min\{\gamma/(\gamma + 4), 1/3\}$.

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 (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 $\mathcal{Z} = \{\mathbf{U}_{t_1, C_1}, \dots, \mathbf{U}_{t_n, C_n}\}$.

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

$$\inf_{\hat{\Sigma} \in \check{\mathcal{F}}} \sup_{\mathcal{P}_2} \mathbb{E}(\|\hat{\Sigma} - \Sigma_u\|_2^2) \gtrsim c_p^2 (n_*^{-1} \log p)^{1-q}$$

provided that $c_p \lesssim n_*^{(1-q)/2} (\log p)^{-(3-q)/2}$, where \mathcal{P}_2 is defined in **Theorem 3**.

Remark 4. (i) If $\{\mathbf{U}_{t_k}\}$ is an independent sequence or L_n -dependent sequence with fixed L_n , **Theorems 3** and **4** indicate that $\hat{\Sigma}_u^{\text{thre}}$ defined as (8) is minimax optimal under the spectral norm loss when we select K as a fixed integer.

(ii) For the general cases with $\varphi < \infty$ and fixed L_n , if we select $K \asymp (\log n_*)^{1/\varphi + \epsilon}$ for some $\epsilon > 0$, **Theorems 3** and **4** indicate that $\hat{\Sigma}_u^{\text{thre}}$ defined as (8) is nearly minimax optimal under the spectral norm with an additional logarithm factor $(\log n_*)^{(1/\varphi + \epsilon)(1-q)/2}$.

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 Σ_u . Another finding from our theoretical analysis is that although the noise $\{\mathbf{U}_{t_1, C_1}, \dots, \mathbf{U}_{t_n, C_n}\}$ are not directly observable, the localized estimator in some scenarios has the (nearly) same accuracy as the one when the noise $\{\mathbf{U}_{t_1, C_1}, \dots, \mathbf{U}_{t_n, C_n}\}$ are observed in the sense of the (nearly) same convergence rates for estimating Σ_u with high-frequency data. From the practical perspective, it can be viewed as a blessing from the high-frequency data with adequate amount of data information locally, so that the statistical properties of the noise can be accurately revealed.

4. The effect of the smallness of the noise

Our results in Section 3 assume that $n_{i,j} \rightarrow \infty$ with T fixed. Empirically, as pointed out in Hansen and Lund (2006), the magnitude of $\sigma_{u,i,j}$ may be small; see also Christensen et al. (2014). To address this issue, we study the second-order property of our estimator concerning its bias. For $\hat{\sigma}_{u,i,j}$ defined as (5), since $\{\mathbf{U}_{t_k}\}_{k=1}^n$ is independent of $(\mathbf{X}_t)_{0 \leq t \leq T}$, we have that

$$\begin{aligned} & \mathbb{E}\{\hat{\sigma}_{u,i,j} \mid (\mathbf{X}_t)_{0 \leq t \leq T}\} - \sigma_{u,i,j} \\ &= - \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \mathbb{E}\left(\frac{U_{i,t_{i,j,k}}}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right)}_{I_2'(i,j)} - \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \mathbb{E}\left(\frac{U_{j,t_{i,j,k}}}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{i,t_{i,j,\ell}}\right)}_{I_3'(i,j)} \\ &+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}})(X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}})}_{\text{II}(i,j)}, \end{aligned} \tag{10}$$

which indicates that the bias in $\hat{\sigma}_{u,i,j}$ includes three parts: $I_2'(i,j)$, $I_3'(i,j)$ and $\text{II}(i,j)$. Proposition 2 in Section 8 shows that $\max_{i,j \in [p]} |\text{II}(i,j)| = O_p\{(Kn_*^{-1} \log p)^{1/2}\}$, but $\mathbb{E}\{\text{II}(i,j)\} \neq 0$, causing a bias of order $O(Kn_*^{-1})$ that summarized in Theorem 5. Under Assumptions 1 and 3, it follows from Davydov’s inequality that $\max_{i,j \in [p]} |I_2'(i,j)| + \max_{i,j \in [p]} |I_3'(i,j)| \lesssim \exp(-C_* L_n^{-\varphi} K^\varphi)$ for some universal constant $C_* > 0$. If $\{\mathbf{U}_{t_k}\}_{k=1}^n$ is an independent sequence or L_n -dependent sequence, we know $\varphi = \infty$ and then $\max_{i,j \in [p]} |I_2'(i,j)| + \max_{i,j \in [p]} |I_3'(i,j)| \lesssim \exp(-\infty) = 0$ with $K > L_n$. If $\varphi < \infty$, with selecting $K = L_n(C_{**} \log n_*)^{1/\varphi}$ for some sufficiently large constant $C_{**} > 0$, $\max_{i,j \in [p]} |I_2'(i,j)| + \max_{i,j \in [p]} |I_3'(i,j)| \lesssim n_*^{-C_{**}}$ will be negligible in comparison to $\text{II}(i,j)$.

Theorem 5. Under Assumptions 2 and 4, if $K = o(n_*)$, it holds that

$$\max_{i,j \in [p]} \left| \mathbb{E}\left\{ \text{II}(i,j) - \frac{2K + \Delta_K}{4n_{i,j}} \int_{t_{i,j,1}}^{t_{i,j,n_{i,j}}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds \right\} \right| \lesssim (Kn_*^{-1})^{3/2}.$$

From (10) and Theorem 5, we have

$$\mathbb{E}(\hat{\sigma}_{u,i,j}) = \sigma_{u,i,j} + \underbrace{\mathbb{E}\left(\frac{2K + \Delta_K}{4n_{i,j}} \int_{t_{i,j,1}}^{t_{i,j,n_{i,j}}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds\right)}_{O(Kn_*^{-1})} + O(K^{3/2} n_*^{-3/2}) \tag{11}$$

provided that $K = L_n(C_{**} \log n_*)^{1/\varphi}$ for some sufficiently large constant $C_{**} > 0$. Since $K = o(n_*)$, the second term on the right-hand side of (11) is asymptotically negligible if $\sigma_{u,i,j}$ is not vanishing; our Theorem 5 implies that it is the leading term in the bias. Impact from the bias on $\hat{\sigma}_{u,i,j}$ could be empirically substantial, especially when $\sigma_{u,i,j}$ is relatively small.

For any $i, j \in [p]$, let $\psi_{i,j} = \int_{t_{i,j,1}}^{t_{i,j,n_{i,j}}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds$. As a remedy, we propose a bias-correction for $\hat{\sigma}_{u,i,j}$ as follows:

$$\hat{\sigma}_{u,i,j}^{bc} = \hat{\sigma}_{u,i,j} - \frac{2K + \Delta_K}{4n_{i,j}} \hat{\psi}_{i,j}, \tag{12}$$

where $\hat{\sigma}_{u,i,j}$ is given in (5), and $\hat{\psi}_{i,j}$ is an estimate of $\psi_{i,j}$. Since $\psi_{i,j}$ is an integrated covariance, it can be estimated by existing approaches, for example, the polarization method (Ait-Sahalia et al., 2010), the two time scales approach (Zhang, 2011), the pre-averaging method (Jacod et al., 2009; Christensen et al., 2010), and the quasi-maximum likelihood approach (Liu and Tang, 2014). Section 6.1.2 gives details for calculating $\hat{\psi}_{i,j}$ by the two time scales approach. Based on $\hat{\sigma}_{u,i,j}^{bc}$ given in (12), we can obtain $\hat{\Sigma}_u^{bc}$ and $\hat{\Sigma}_u^{bc, \text{thre}}$, the bias-corrected version of $\hat{\Sigma}_u$ defined as (6) and $\hat{\Sigma}_u^{\text{thre}}$ defined as (8), respectively, by replacing $\hat{\sigma}_{u,i,j}$ by $\hat{\sigma}_{u,i,j}^{bc}$. Theorem 6 indicates that $\hat{\Sigma}_u^{bc}$ and $\hat{\Sigma}_u^{bc, \text{thre}}$ share the same convergence rates of $\hat{\Sigma}_u$ and $\hat{\Sigma}_u^{\text{thre}}$, respectively.

Theorem 6. Assume $\max_{i,j \in [p]} |\hat{\psi}_{i,j}| = O_p(\log p)$. The following two assertions are satisfied:

- (i) Under the conditions of [Theorem 1](#), $|\hat{\Sigma}_u^{bc} - \Sigma_u|_\infty = O_p\{(Kn_*^{-1} \log p)^{1/2}\}$.
- (ii) Under the conditions of [Theorem 3](#), $\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2 = O_p\{c_p(Kn_*^{-1} \log p)^{(1-q)/2}\}$ for any covariance matrix $\Sigma_u \in \mathcal{H}(q, c_p, M)$.

5. Impact from jumps and the robustness of our methods

We now consider the setting with jumps in the underlying process \mathbf{X}_t . Assume $\mathbf{X}_t = (X_{1,t}, \dots, X_{p,t})^\top$ satisfies the following model:

$$dX_{i,t} = \mu_{i,t} dt + \sigma_{i,t} dB_{i,t} + J_{i,t} dM_{i,t} \quad \text{and} \quad \mathbb{E}(dB_{i,t} \cdot dB_{j,t}) = \rho_{i,j,t} dt, \tag{13}$$

where $\mu_{i,t}$'s, $\sigma_{i,t}$'s, $B_{i,t}$'s and $\rho_{i,j,t}$'s are same as those in (1), $J_{i,t}$'s are the jump sizes, and $M_{i,t}$'s are counting processes. Our analysis reveals that our estimators proposed in Section 2.2 for Σ_u are reasonably robust against jumps. In our theoretical analysis, we impose [Assumptions 6](#) and [7](#) on the counting process $M_{i,t}$ and the jump size $J_{i,t}$, respectively.

Assumption 6. Let $(\Delta M_{i,\cdot})_{i,j} = M_{i,t_{i,j},n_{i,j}} - M_{i,t_{i,j},1}$ for any $i, j \in [p]$. There exist $\lambda_1, \dots, \lambda_p > 0$ and some universal constants $C_{10} > 0, C_{11} > 0$ such that (i) $\mathbb{E}\{\exp(\theta[(\Delta M_{i,\cdot})_{i,j} - \mathbb{E}\{(\Delta M_{i,\cdot})_{i,j}\}])\} \leq \exp\{\lambda_i(t_{i,j},n_{i,j} - t_{i,j},1)\theta^2\}$ for any $|\theta| \leq C_{10}^{-1}$ and $i \in [p]$; (ii) $\mathbb{E}\{(\Delta M_{i,\cdot})_{i,j}\} \leq C_{11}\lambda_i(t_{i,j},n_{i,j} - t_{i,j},1)$ for any $i \in [p]$.

Assumption 7. There exist some universal constants $\iota > 0, C_{12} > 1$ and $C_{13} > 0$ such that $\mathbb{P}(\sup_{0 \leq t \leq T} |J_{i,t}| > u) \leq C_{12} \exp(-C_{13}u^\iota)$ for any $i \in [p]$ and $u > 0$.

If $M_{i,t}$ is a Poisson process with intensity $\lambda_i > 0$, then $(\Delta M_{i,\cdot})_{i,j}$ follows Poisson distribution with parameter $\lambda_i(t_{i,j},n_{i,j} - t_{i,j},1)$ and [Assumption 6](#) holds for $C_{10} = C_{11} = 1$. [Assumption 7](#) controls the tail behavior of the random jump size $\sup_{0 \leq t \leq T} J_{i,t}$. If the jump size $J_{i,t}$ is bounded from above uniformly for $i \in [p]$ and $t \in [0, T]$, we can select $\iota = \infty$ in [Assumption 7](#).

Recall $S_{i,j,k} = \{t_{i,j,\ell} \in \mathcal{G}_i \cap \mathcal{G}_j : K \leq |\ell - k| \leq K + \Delta_K\}$ for some integers $K \geq 1$ and $\Delta_K \geq 0$, and $N_{i,j,k} = |S_{i,j,k}|$. For any $i, j \in [p]$, define

$$\varpi_{i,j} = \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{i,s} dM_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{j,s} dM_{j,s} \right). \tag{14}$$

Let $\Omega = (\varpi_{i,j})_{p \times p}$ and

$$\hat{\Sigma}_u^{\text{jump}} = (\hat{\sigma}_{u,i,j}^{\text{jump}})_{p \times p} = \hat{\Sigma}_u - \Omega,$$

where $\hat{\Sigma}_u$ is defined as (6). Analogous to (8), we define the thresholding version of $\hat{\Sigma}_u^{\text{jump}}$ as

$$\hat{\Sigma}_u^{\text{jump, thre}} = [\hat{\sigma}_{u,i,j}^{\text{jump}} I\{|\hat{\sigma}_{u,i,j}^{\text{jump}}| \geq \beta(Kn_*^{-1} \log p)^{1/2}\}]_{p \times p}, \tag{15}$$

where $\beta > 0$ is a fixed constant for the thresholding level.

Our theory has two parts. As the first part, parallel to [Theorems 1](#) and [3](#), we have the next theorem for the convergence rates of $\hat{\Sigma}_u^{\text{jump}}$ and $\hat{\Sigma}_u^{\text{jump, thre}}$.

Theorem 7. Assume [Assumptions 1–7](#) hold. Let $\lambda_* = \max_{i \in [p]} \lambda_i$ and $K > CL_n$ for some constant $C \geq 1$. If $\lambda_*^2(n_*^{-1}K)^{\iota/(2\iota+2)}(\log p)^{-1} = o(1)$, $\log p = o[\min\{(n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}, (n_*K^{-1})^{\tilde{\chi}}\}]$ and $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, where n_* is specified in (4) and $\tilde{\chi} = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2), \gamma/(\gamma + 4), 1/3\}$, then the following two assertions are satisfied:

(i) Let \mathcal{P}_3 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 [Assumption 3](#), $\mathbf{X}_t = (X_{1,t}, \dots, X_{p,t})^\top$ follows model (13) with each $\mu_{i,t}, \sigma_{i,t}, M_{i,t}$ and $J_{i,t}$ satisfying [Assumptions 4–7](#), and the grids of time points $\{\mathcal{G}_i\}_{i=1}^p$ satisfy [Assumption 2](#). It holds that

$$\sup_{\mathcal{P}_3} \mathbb{E}(|\hat{\Sigma}_u^{\text{jump}} - \Sigma_u|_\infty) \lesssim (Kn_*^{-1} \log p)^{1/2}.$$

(ii) Let \mathcal{P}_4 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 [Assumption 3](#) with the covariance matrix $\Sigma_u \in \mathcal{H}(q, c_p, M)$, $\mathbf{X}_t = (X_{1,t}, \dots, X_{p,t})^\top$ follows model (13) with each $\mu_{i,t}, \sigma_{i,t}, M_{i,t}$ and $J_{i,t}$ satisfying [Assumptions 4–7](#), and the grids of time points $\{\mathcal{G}_i\}_{i=1}^p$ satisfy [Assumption 2](#). With sufficiently large constant $\beta > 0$ in (15), it holds that

$$\sup_{\mathcal{P}_4} \mathbb{E}(\|\hat{\Sigma}_u^{\text{jump, thre}} - \Sigma_u\|_2^2) \lesssim c_p^2(Kn_*^{-1} \log p)^{1-q}.$$

Theorems 7(i) and **7(ii)** can be viewed as the generalization of **Theorems 1** and **3**, respectively. If there are no jumps in \mathbf{X}_t , we have $J_{i,t} \equiv 0$ for any $i \in [p]$ and $t \in [0, T]$, and then $\widehat{\Sigma}_u^{\text{jump}} = \widehat{\Sigma}_u$ and $\widehat{\Sigma}_u^{\text{jump,thre}} = \widehat{\Sigma}_u^{\text{thre}}$. In this scenario, we can set $\lambda_i = 0$ for any $i \in [p]$ and $\iota = \infty$ in **Assumptions 6** and **7**, respectively, which implies $\lambda_*^2(n_*^{-1}K)^{\iota/(2\iota+2)}(\log p)^{-1} = o(1)$ holds automatically and $\check{\chi} = \min\{\gamma/(\gamma + 4), 1/3\}$. Hence, the results of **Theorems 7(i)** and **7(ii)** in this scenario are identical to **Theorems 1** and **3**, respectively.

In the second part of our theory, we establish the properties of $\widehat{\Sigma}_u$ itself. Since $\widehat{\Sigma}_u^{\text{jump}} = \widehat{\Sigma}_u - \Omega$ where $\Omega = (\varpi_{i,j})_{p \times p}$ with $\varpi_{i,j}$ defined as (14), the gap between $\widehat{\Sigma}_u$ and $\widehat{\Sigma}_u^{\text{jump}}$ is seen determined by an extra 'bias'. We assume the following condition for controlling the tail probability of $\varpi_{i,j}$.

Assumption 8. Each $M_{i,t}$ has independent increments. Let $(\Delta M_{i,\cdot})_{i,j}^{(k)} = M_{i,t_{i,j,k+1}} - M_{i,t_{i,j,k}}$ for any $i, j \in [p]$ and $k \in [n_{i,j} - 1]$. There exist $\lambda_1, \dots, \lambda_p > 0$ and some universal constants $C_{10} > 0, C_{11} > 0$ such that (i) $\mathbb{E}\{\exp(\theta[(\Delta M_{i,\cdot})_{i,j}^{(k)} - \mathbb{E}\{(\Delta M_{i,\cdot})_{i,j}^{(k)}\}])\} \leq \exp\{\lambda_i(t_{i,j,k+1} - t_{i,j,k})\theta^2\}$ for any $|\theta| \leq C_{10}^{-1}$ and $i \in [p]$; (ii) $\mathbb{E}\{(\Delta M_{i,\cdot})_{i,j}^{(k)}\} \leq C_{11}\lambda_i(t_{i,j,k+1} - t_{i,j,k})$ for any $i \in [p]$.

Assumption 6 holds automatically under **Assumption 8**. If $M_{i,t}$ is a Poisson process with intensity $\lambda_i > 0$, then $(\Delta M_{i,\cdot})_{i,j}^{(k)}$ follows the Poisson distribution with parameter $\lambda_i(t_{i,j,k+1} - t_{i,j,k})$ and **Assumption 8** holds for $C_{10} = C_{11} = 1$. Given **Assumptions 7** and **8**, we have the following theorem for the tail probability of $\varpi_{i,j}$.

Theorem 8. Let $\lambda_* = \max_{i \in [p]} \lambda_i$. Under **Assumptions 7** and **8**, if $K(\log n_*)^{(3\iota+4)/\iota} = o(n_*)$ and $n_*^{-1}K\lambda_* = o(1)$, it holds that

$$\max_{i,j \in [p]} \mathbb{P}\{|\varpi_{i,j}| \geq v\} \lesssim \exp\{-C(n_*K^{-1})^{(2\iota+2)/(3\iota+4)}v\} + \exp\{-C(n_*K^{-1})^{\iota/(3\iota+4)}\}$$

for any $v \gg (n_*^{-1}K)^{(2\iota+2)/(3\iota+4)}\lambda_*$, where ι is specified in **Assumption 7**. Furthermore, it holds that $\max_{i,j \in [p]} \mathbb{E}\{|\varpi_{i,j}|^m\} \lesssim 1$ for any fixed positive integer m .

Theorem 8 implies $|\Omega|_\infty = O_p\{(Kn_*^{-1})^{(2\iota+2)/(3\iota+4)} \log p\} = o_p\{(Kn_*^{-1} \log p)^{1/2}\}$ if $\log p = o\{(n_*K^{-1})^{\iota/(3\iota+4)}\}$, which leads to the robustness of our proposed estimators against possible jumps in the underlying process \mathbf{X}_t , as established in the following theorem.

Theorem 9. Let $\lambda_* = \max_{i \in [p]} \lambda_i$ and $K > CL_n$ for some constant $C \geq 1$. Under **Assumptions 1–5** and **7–8**, if $\lambda_*^2(n_*^{-1}K)^{\iota/(3\iota+4)}(\log p)^{-1} = o(1)$, $\log p = o[\min\{(n_*L_n^{-2}K)^\varphi/(3\varphi+2), (n_*K^{-1})^{\chi^*}\}]$ and $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, where n_* is specified in (4) and $\chi^* = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(3\iota + 4), \gamma/(\gamma + 4)\}$, then the following two assertions are satisfied:

- (i) $|\widehat{\Sigma}_u - \Sigma_u|_\infty = O_p\{(Kn_*^{-1} \log p)^{1/2}\}$.
- (ii) With sufficiently large constant $\beta > 0$ in (8), $\|\widehat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2 = O_p\{c_p(Kn_*^{-1} \log p)^{(1-q)/2}\}$ for any covariance matrix $\Sigma_u \in \mathcal{H}(q, c_p, M)$.

Specifically, **Theorem 9** implies that if the intensity parameter λ_* of the counting processes diverges no faster than $(n_*K^{-1})^{\iota/(6\iota+8)}(\log p)^{1/2}$, the convergence rates of $\widehat{\Sigma}_u$ and $\widehat{\Sigma}_u^{\text{thre}}$ are the same as those in **Theorems 1** and **3**, even there are jumps in the processes.

6. Numerical studies

6.1. Simulations

6.1.1. Data generating procedure

We set $(\rho_{i,j})_{p \times p} = \{\text{diag}(\mathbf{A}^{\otimes 2})\}^{-1/2} \mathbf{A}^{\otimes 2} \{\text{diag}(\mathbf{A}^{\otimes 2})\}^{-1/2}$, where $\mathbf{A} = (a_{i,j})_{p \times p}$ with $a_{i,j} = (-0.8)^{|i-j|}I(i \geq j)$. For $\mathbf{X}_t = (X_{1,t}, \dots, X_{p,t})^\top$, we generated each $X_{i,t}$ from the following stochastic volatility model:

$$dX_{i,t} = \sigma_{i,t} dB_{i,t} + J_{i,t} dM_{i,t}, \quad d\sigma_{i,t}^2 = \kappa(\bar{\sigma}^2 - \sigma_{i,t}^2) dt + s\sigma_{i,t} dW_{i,t}, \tag{16}$$

where $M_{1,t}, \dots, M_{p,t}$ are independent Poisson processes with intensity λ_j , and $B_{1,t}, \dots, B_{p,t}, W_{1,t}, \dots, W_{p,t}$ are univariate standard Brownian motions such that (i) $W_{1,t}, \dots, W_{p,t}$ are p independent Brownian motions, and (ii) $\mathbb{E}(dB_{i,t} \cdot dW_{j,t}) = -0.3I(i = j)dt$ and $\mathbb{E}(dB_{i,t} \cdot dB_{j,t}) = \rho_{i,j} dt$. We considered two settings – with or without jumps: (i) $J_{1,t}, \dots, J_{p,t} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_j^2)$ which are independent of $M_{1,t}, \dots, M_{p,t}$, and (ii) $J_{i,t} \equiv 0$ for $i \in [p]$ and t . We set $(\kappa, s, \bar{\sigma}^2, \sigma_j^2, \lambda_j) = (5, 0.5, 0.1, 0.015^2, 5)$, the same as that in the numerical studies of **Ait-Sahalia et al. (2013)** that mimics the empirical features of financial data (**Ait-Sahalia and Kimmel, 2007**). This setting is reasonable; comparable settings are found in existing studies (**Ait-Sahalia and Yu, 2009; Ait-Sahalia et al., 2010; Liu and Tang, 2014; Ait-Sahalia and Xiu, 2017**). The initial observations of $\sigma_{i,t}^2$ ($i \in [p]$) were generated from a Gamma distribution $\Gamma(2\kappa\bar{\sigma}^2/s^2, s^2/(2\kappa))$. In our simulation, we set $p \in \{50, 100, 200\}$.

We took $t \in [0, T]$ with $T = 1/252$; here 1 unit of t means one year, so $T = 1/252$ is corresponding to a trading day. We first generated high-frequency data available at each second in a 6.5-hour period; this setting results

in $60 \times 60 \times 6.5 = 23400$ observations. By letting $\tilde{t}_k = k/(252 \times 23400)$, we generated $\mathbf{Y}_{\tilde{t}_k} = \mathbf{X}_{\tilde{t}_k} + \mathbf{U}_{\tilde{t}_k}$ with $(\mathbf{X}_t)_{t \in [0, T]}$ from (16), and each element of $\mathbf{U}_{\tilde{t}_k} = (U_{1, \tilde{t}_k}, \dots, U_{p, \tilde{t}_k})^\top$ from a stationary GARCH(1,1) model:

$$U_{i, \tilde{t}_k} = \tilde{\sigma}_{i, \tilde{t}_k} \eta_{i, \tilde{t}_k}, \quad \tilde{\sigma}_{i, \tilde{t}_k}^2 = 0.1\sigma_e^2 + 0.1\tilde{\sigma}_{i, \tilde{t}_{k-1}}^2 \eta_{i, \tilde{t}_{k-1}}^2 + 0.8\tilde{\sigma}_{i, \tilde{t}_{k-1}}^2,$$

where $\eta_{\tilde{t}_k} = (\eta_{1, \tilde{t}_k}, \dots, \eta_{p, \tilde{t}_k})^\top$ is independently generated from $N(\mathbf{0}, \mathbf{R})$; the settings of \mathbf{R} will be described later. In this model, upon observing $\mathbb{E}(\tilde{\sigma}_{i, \tilde{t}_k}^2) = \sigma_e^2$ for each $i \in [p]$, we considered different settings for the signal-to-noise ratio $\sigma_e^2/(T^{-1} \int_0^T \sigma_{i,t}^2 dt)$ by varying σ_e^2 . Since the signal-to-noise ratio can be approximated by $252\sigma_e^2/\bar{\sigma}^2$ with $\bar{\sigma}^2$ in (16), we specified two selections of σ_e^2 : (i) $\sigma_e^2 = 0.005^2$ such that $252\sigma_e^2/\bar{\sigma}^2 = 0.063$, and (ii) $\sigma_e^2 = 0.001^2$ such that $252\sigma_e^2/\bar{\sigma}^2 = 0.00252$. In Part S1 of the supplementary material, we have also investigated the finite-sample performance of the proposed estimators when $\mathbf{U}_{\tilde{t}_1}, \dots, \mathbf{U}_{\tilde{t}_{23400}} \stackrel{i.i.d.}{\sim} N(\mathbf{0}, \sigma_e^2 \mathbf{R})$ with $\sigma_e^2 = 0.005^2$ and 0.001^2 .

We studied the following three models for $\mathbf{R} = (r_{i,j})_{p \times p}$ that controls the correlations:

Model 1: \mathbf{R} is a banded matrix, where $r_{i,i} = 1, r_{i+1,i} = r_{i,i+1} = 0.6, r_{i+2,i} = r_{i,i+2} = 0.3$, and $r_{i,j} = 0$ for $|i - j| \geq 3$.

Model 2: $\mathbf{R} = \{\text{diag}(\mathbf{R}^*)\}^{-1/2} \mathbf{R}^* \{\text{diag}(\mathbf{R}^*)\}^{-1/2}$, where $\mathbf{R}^* = \tilde{\mathbf{R}} + \{|\lambda_{\min}(\tilde{\mathbf{R}})| + 0.05\} \mathbf{I}_p$, \mathbf{I}_p is the identity matrix of order p , $\lambda_{\min}(\tilde{\mathbf{R}})$ is the smallest eigenvalue of $\tilde{\mathbf{R}}$, and $\tilde{\mathbf{R}} = (\tilde{r}_{i,j})_{p \times p}$ satisfies that $\tilde{r}_{i,j} = w_{i,j} b_{i,j}$, $w_{i,j}$'s are independently generated from the uniform distribution $U(0.4, 0.8)$, $b_{i,j}$'s are independently generated from the Bernoulli distribution with successful probability 0.04.

Model 3: \mathbf{R} is a bandable matrix with $r_{i,j} = 0.6^{|i-j|}$.

We considered both synchronous and asynchronous high-frequency data in our simulation. To model the synchronous data setting, we took $\{\mathbf{Y}_{\tilde{t}_{k\Delta}}\}_{k=1}^{\lfloor 23400/\Delta \rfloor}$ as the observed data where $\lfloor x \rfloor$ is the floor function; by varying Δ , we simulated data sets of different sizes: larger Δ means fewer observations. Then the time points where we observed the noisy data are $t_k = \tilde{t}_{k\Delta}$ with $k = 1, \dots, \lfloor 23400/\Delta \rfloor$. In our numerical studies, we set $\Delta \in \{1, 2, 3\}$. To model the asynchronous data setting, for each $i \in [p]$, we applied the Poisson process sampling scheme with intensity λ to $\{\tilde{t}_k\}_{k=1}^{23400}$ for generating $\mathcal{G}_i = \{t_{i,1}, \dots, t_{i,n_i}\}$, the grid of time points at which we actually observed $Y_{i,t}$. The Poisson process sampling schemes for different i 's are independent. Based on this setting for asynchronous data, on average there were $\lfloor 23400/\lambda \rfloor$ observations for each $Y_{i,t}$. We selected $\lambda \in \{1, 2, 3\}$ in our simulation.

6.1.2. Implementation of bias-correction

To obtain the bias-corrected estimator $\hat{\sigma}_{u,i,j}^{bc}$ in (12), we need to calculate $\hat{\psi}_{i,j}$, the estimate of the integrated covariance $\psi_{i,j}$. In the simulation, we applied the two time scales approach (Zhang, 2011) to estimate $\psi_{i,j}$. Recall $\mathcal{G}_i = \{t_{i,1}, \dots, t_{i,n_i}\}$ is the grid of time points we observed $Y_{i,t}$. For given $i, j \in [p]$, we first used the refresh time procedure (Barndorff-Nielsen et al., 2011) to synchronize the data if $\mathcal{G}_i \neq \mathcal{G}_j$. More specifically, let the first refresh time point be $v_1 = \max\{t_{i,1}, t_{j,1}\}$, and then define the other refresh time points v_{l+1} with $l \geq 1$ as $v_{l+1} = \max\{\min\{t \in \mathcal{G}_i : t > v_l\}, \min\{t \in \mathcal{G}_j : t > v_l\}\}$ iteratively. Denoted by $n_{i,j}^*$ the resulting refresh time points for \mathcal{G}_i and \mathcal{G}_j , and write $t_{i,l}^* = \max\{t \in \mathcal{G}_i : t \leq v_l\}$ and $t_{j,l}^* = \max\{t \in \mathcal{G}_j : t \leq v_l\}$ for each $l \in [n_{i,j}^*]$. If $\mathcal{G}_i = \mathcal{G}_j$, the refresh time points based on above procedure are identical to $t_{i,1}, \dots, t_{i,n_i}$, and thus the associated $t_{i,l}^* = t_{j,l}^* = t_{i,l}$. For given positive integers δ_1 and δ_2 , the two time scales estimator for $\psi_{i,j}$ is given by

$$\hat{\psi}_{i,j} = [Y_i, Y_j]^{(\delta_1)} - \frac{\delta_2(n_{i,j}^* - \delta_1 + 1)}{\delta_1(n_{i,j}^* - \delta_2 + 1)} [Y_i, Y_j]^{(\delta_2)}, \tag{17}$$

where $[Y_i, Y_j]^{(\delta)} = \delta^{-1} \sum_{l=\delta+1}^{n_{i,j}^*} (Y_{i,t_{i,l}^*} - Y_{i,t_{i,l-\delta}^*})(Y_{j,t_{j,l}^*} - Y_{j,t_{j,l-\delta}^*})$ for any positive integer δ . Following Ait-Sahalia and Yu (2009), we set $(\delta_1, \delta_2) = (25, 1)$ in our simulation.

6.1.3. Selections of (K, Δ_K) and the thresholding level

To obtain $\hat{\sigma}_{u,i,j}$ defined as (5) in practice, we need to select the tuning parameters K and Δ_K . If each univariate sequence $\{U_{i,t_k}\}_{k=1}^n$ is α -mixing with exponentially decaying α -mixing coefficients,² with the independence assumption imposed on the p sequences $\{U_{1,t_k}\}_{k=1}^n, \dots, \{U_{p,t_k}\}_{k=1}^n$, Theorem 5.1 of Bradley (2005) indicates that $\alpha_n(m)$ defined in (3) satisfies $\alpha_n(m) \leq p \exp(-cm)$ for some universal constant $c > 0$, which provides a rough upper bound for $\alpha_n(m)$. Hence, Assumption 1 holds for $\varphi = 1$ and $L_n = \tilde{c} \log p$ for some sufficiently small constant $\tilde{c} > 0$. Our theoretical results require $K^{-\varphi} L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$ and $K > CL_n$ for some constant $C \geq 1$. To match these requirements, when we estimate $\sigma_{u,i,j}$, we can select $K = \tilde{c}(\log p)(\log n_{i,j})(\log \log n_{i,j})$ for some small constant $\tilde{c} > 0$. In our simulation, we have tried $\tilde{c} \in [0.03, 0.07]$ and the associated results are similar. We suggest to select $\tilde{c} = 0.05$ in practice. Since we use $(Y_{i,t_{i,j,\ell}} - Y_{i,t_{i,j,k}}, Y_{j,t_{j,\ell}} - Y_{j,t_{j,k}})$ to approximate $(U_{i,t_{i,j,\ell}} - U_{i,t_{i,j,k}}, U_{j,t_{j,\ell}} - U_{j,t_{j,k}})$ for $K \leq |\ell - k| \leq K + \Delta_K$, the bias

² Such requirement can be easily satisfied in most commonly used univariate time series models. See our discussion below (3).

issue caused by $X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}}$ and $X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}}$ will impact the performance of our estimators. Notice that a smaller Δ_K results in a smaller bias. We need to select Δ_K as some small positive integers. Table 1 shows that (i) the estimators with $\Delta_K \in \{1, 2, 3\}$ perform quite well, and (ii) the estimators with $\Delta_K = 1$ work best in most cases and perform quite close to the best ones in other cases. This verifies our claim that Δ_K should be selected as some small integers. We suggest to select $\Delta_K \in \{1, 2, 3\}$ in practice.

Based on $\widehat{\Sigma}_u = (\widehat{\sigma}_{u,i,j})_{p \times p}$ and $\widehat{\Sigma}_u^{bc} = (\widehat{\sigma}_{u,i,j}^{bc})_{p \times p}$, to derive their thresholding version $\widehat{\Sigma}_u^{thre}$ and $\widehat{\Sigma}_u^{bc,thre}$, we need to determine the thresholding level. Our theoretical analysis shows that the thresholding level should have the order $(Kn_*^{-1} \log p)^{1/2}$. Notice that $\widehat{\sigma}_{u,i,j} = n_{i,j}^{-1} \sum_{k=1}^{n_{i,j}} \zeta_{i,j,k}$ with $\zeta_{i,j,k} = (2N_{i,j,k})^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (Y_{i,t_{i,j,\ell}} - Y_{i,t_{i,j,k}})(Y_{j,t_{i,j,\ell}} - Y_{j,t_{i,j,k}})$. Since $\text{Var}(\widehat{\sigma}_{u,i,j})$ has the order Kn_*^{-1} , the long-run variance of the sequence $\{\zeta_{i,j,k}\}_{k=1}^{n_{i,j}}$ has the order K . To incorporate the heterogeneity of the estimators $\widehat{\sigma}_{u,i,j}$, we implemented the thresholding estimators in practice as

$$\begin{aligned} \widehat{\Sigma}_u^{thre} &= (\widehat{\sigma}_{u,i,j}^{thre})_{p \times p} = [\widehat{\sigma}_{u,i,j} I\{\widehat{\sigma}_{u,i,j} \geq \beta_* (\widehat{\theta}_{i,j} n_{i,j}^{-1} \log p)^{1/2}\}]_{p \times p}, \\ \widehat{\Sigma}_u^{bc,thre} &= (\widehat{\sigma}_{u,i,j}^{bc,thre})_{p \times p} = [\widehat{\sigma}_{u,i,j}^{bc} I\{|\widehat{\sigma}_{u,i,j}^{bc}| \geq \beta_* (\widehat{\theta}_{i,j} n_{i,j}^{-1} \log p)^{1/2}\}]_{p \times p}, \end{aligned} \tag{18}$$

where $\beta_* > 0$ is a constant, and $\widehat{\theta}_{i,j}$ is an estimate for the long-run variance of the sequence $\{\zeta_{i,j,k}\}_{k=1}^{n_{i,j}}$. Write $\bar{\zeta}_{i,j} = n_{i,j}^{-1} \sum_{k=1}^{n_{i,j}} \zeta_{i,j,k}$. We chose $\widehat{\theta}_{i,j}$ in (18) as

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

where $\kappa(\cdot)$ is a symmetric kernel function, h is the bandwidth, $\widehat{H}_{i,j}(\ell) = n_{i,j}^{-1} \sum_{k=\ell+1}^{n_{i,j}} (\zeta_{i,j,k} - \bar{\zeta}_{i,j})(\zeta_{i,j,k-\ell} - \bar{\zeta}_{i,j})$ for $\ell \geq 0$ and $\widehat{H}_{i,j}(\ell) = n_{i,j}^{-1} \sum_{k=-\ell+1}^{n_{i,j}} (\zeta_{i,j,k+\ell} - \bar{\zeta}_{i,j})(\zeta_{i,j,k} - \bar{\zeta}_{i,j})$ otherwise. Andrews (1991) suggested the quadratic spectral kernel

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

with optimal bandwidth $h = 1.3221\{4n_{i,j}\widehat{\vartheta}_{i,j}^2(1 - \widehat{\vartheta}_{i,j})^{-4}\}^{1/5}$, where $\widehat{\vartheta}_{i,j}$ is the estimated autoregressive coefficient in the fitted AR(1) model for the sequence $\{\zeta_{i,j,k}\}_{k=1}^{n_{i,j}}$. In our simulation, we have tried $\beta_* \in [1.75, 2.25]$ and the associated results are similar. We suggest to select $\beta_* = 2$ in practice.

6.1.4. Simulation results

For given estimator $\widetilde{\Sigma}$, we evaluated its relative estimation error $\|\widetilde{\Sigma} - \Sigma_u\|_2 / \|\Sigma_u\|_2$ in different settings. Table 1 summarizes the averages of the relative estimation errors based on 1000 repetitions. We have several observations. First, we find that in general, $\widehat{\Sigma}_u^{bc,thre}$ performs quite well for all cases with satisfactorily small relative estimation errors compared with $\widehat{\Sigma}_u^{thre}$. Further, $\widehat{\Sigma}_u^{thre}$ performs quite well when $252\sigma_e^2 / \bar{\sigma}^2 = 0.063$ but poorly when $252\sigma_e^2 / \bar{\sigma}^2 = 0.00252$. This suggests that when the noise is quite small, the bias-correction is necessary. Second, as the dimension p increases, the relative estimation errors worsen a bit, but at a very slow pace growing with p . This demonstrates the promising performance of the thresholding method for handling high-dimensional covariance estimations. Third, as the sampling frequency becomes higher (smaller Δ or λ), the performance is 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 with asynchronous data when Δ and λ are the same. Fourth, we find that the differences are small among the performances with different Δ_K , especially when the data are synchronous. Fifth, we find that the empirical performance of the proposed estimators is robust to jumps, confirming our finding in Theorem 9.

In addition, for given estimator $\widetilde{\Sigma} = (\widetilde{\sigma}_{i,j})_{p \times p}$, we also calculated in Tables 2 and 3 the true positive rate (TPR) and the false positive rate (FPR) defined as

$$\begin{aligned} \text{TPR} &= \frac{| \{(i, j) : \widetilde{\sigma}_{i,j} \neq 0 \text{ and } \sigma_{u,i,j} \neq 0 \} |}{| \{(i, j) : \sigma_{u,i,j} \neq 0 \} |}, \\ \text{FPR} &= \frac{| \{(i, j) : \widetilde{\sigma}_{i,j} \neq 0 \text{ and } \sigma_{u,i,j} = 0 \} |}{| \{(i, j) : \sigma_{u,i,j} = 0 \} |}. \end{aligned}$$

Since the covariance matrix considered in Model 3 has no exact zero element, we omit reporting the TPR and FPR in this case. Results in Tables 2 and 3 show that the TPRs of our proposed estimators for all cases are equal to 1 or quite close to 1, and the FPRs for all cases are almost 0. This indicates that our proposed thresholding method can recover the non-zero elements of the covariance matrix very accurately. From the results in Table 3 when the data are asynchronous with $\lambda = 2$ and 3, we find that $\widehat{\Sigma}_u^{bc,thre}$ performs a bit better than $\widehat{\Sigma}_u^{thre}$, and both $\widehat{\Sigma}_u^{bc,thre}$ and $\widehat{\Sigma}_u^{thre}$ have lower TPRs when $252\sigma_e^2 / \bar{\sigma}^2 = 0.00252$, which is reasonable as the signal-to-noise ratio in terms of estimating the covariance matrix of noises is lower in this case. For the FPRs, we find that there is no big difference between the FPRs of $\widehat{\Sigma}_u^{bc,thre}$ with

Table 1

Averages of the relative estimation errors ($\times 100$) for the proposed estimators when jumps exist and do not exist (in parentheses) based on 1000 repetitions.

Synchronous data			Model 1			Model 2			Model 3			
$252\sigma_\epsilon^2/\bar{\sigma}^2$	p	Estimators	Δ_K	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$
0.063	50	$\hat{\Sigma}_u^{thre}$	1	4.5(4.5)	4.0(4.0)	3.2(3.2)	5.2(5.5)	4.6(4.9)	3.3(3.4)	6.3(6.3)	5.3(5.3)	4.3(4.3)
			2	4.5(4.5)	4.0(4.0)	3.2(3.2)	5.3(5.6)	4.7(4.9)	3.3(3.5)	6.1(6.0)	5.2(5.1)	4.2(4.2)
			3	4.6(4.6)	4.1(4.1)	3.2(3.3)	5.6(5.9)	4.8(5.1)	3.4(3.6)	5.9(5.9)	5.0(5.0)	4.1(4.1)
		$\hat{\Sigma}_u^{bc,thre}$	1	4.6(4.6)	4.1(4.2)	3.3(3.3)	5.3(5.5)	4.7(4.9)	3.4(3.5)	6.4(6.4)	5.4(5.4)	4.4(4.3)
			2	4.6(4.6)	4.2(4.2)	3.3(3.3)	5.3(5.6)	4.7(5.0)	3.4(3.6)	6.1(6.1)	5.2(5.2)	4.2(4.2)
			3	4.7(4.7)	4.2(4.3)	3.3(3.4)	5.6(5.9)	4.9(5.2)	3.5(3.7)	6.0(6.0)	5.1(5.1)	4.1(4.1)
	100	$\hat{\Sigma}_u^{thre}$	1	5.2(5.2)	4.4(4.4)	3.6(3.6)	5.7(5.9)	4.6(4.7)	3.5(3.5)	7.0(6.9)	6.0(6.0)	4.8(4.7)
			2	5.2(5.2)	4.4(4.4)	3.6(3.6)	5.8(5.9)	4.6(4.7)	3.5(3.6)	6.7(6.7)	5.8(5.8)	4.6(4.6)
			3	5.3(5.3)	4.5(4.5)	3.6(3.7)	6.0(6.1)	4.8(4.9)	3.6(3.7)	6.6(6.5)	5.7(5.7)	4.6(4.5)
		$\hat{\Sigma}_u^{bc,thre}$	1	5.3(5.4)	4.5(4.5)	3.7(3.7)	5.8(6.0)	4.7(4.8)	3.6(3.7)	7.1(7.0)	6.1(6.1)	4.8(4.8)
			2	5.4(5.4)	4.5(4.6)	3.7(3.7)	5.9(6.1)	4.8(4.9)	3.7(3.7)	6.9(6.8)	5.9(5.9)	4.7(4.6)
			3	5.5(5.5)	4.6(4.7)	3.8(3.8)	6.1(6.3)	5.0(5.1)	3.8(3.9)	6.8(6.7)	5.8(5.8)	4.6(4.6)
200	$\hat{\Sigma}_u^{thre}$	1	6.0(6.0)	5.0(5.0)	4.0(4.0)	5.5(5.5)	4.4(4.5)	3.3(3.4)	7.5(7.6)	6.5(6.5)	5.2(5.1)	
		2	5.9(5.9)	5.0(5.0)	4.0(4.0)	5.4(5.5)	4.4(4.5)	3.3(3.3)	7.3(7.3)	6.3(6.3)	5.0(5.0)	
		3	5.9(5.9)	5.0(5.0)	4.0(4.0)	5.4(5.5)	4.5(4.5)	3.3(3.3)	7.2(7.2)	6.3(6.2)	5.0(4.9)	
	$\hat{\Sigma}_u^{bc,thre}$	1	6.2(6.2)	5.2(5.2)	4.2(4.2)	5.7(5.7)	4.6(4.7)	3.6(3.6)	7.7(7.8)	6.6(6.6)	5.3(5.2)	
		2	6.3(6.3)	5.3(5.3)	4.2(4.2)	5.7(5.7)	4.7(4.7)	3.6(3.6)	7.5(7.6)	6.5(6.5)	5.1(5.1)	
		3	6.3(6.3)	5.4(5.4)	4.3(4.3)	5.8(5.8)	4.8(4.8)	3.6(3.7)	7.5(7.5)	6.4(6.4)	5.1(5.1)	
0.00252	50	$\hat{\Sigma}_u^{thre}$	1	23.5(18.6)	21.0(16.6)	9.9(8.0)	40.6(34.8)	36.4(30.9)	16.5(14.2)	19.9(16.3)	17.1(14.1)	9.0(7.5)
			2	28.2(22.1)	24.3(19.0)	11.5(9.1)	49.0(41.6)	42.1(35.7)	19.4(16.6)	22.7(18.6)	19.0(15.7)	10.0(8.3)
			3	33.0(25.9)	27.5(21.6)	13.1(10.4)	57.4(49.0)	47.6(40.7)	22.4(19.0)	25.5(20.9)	20.9(17.3)	10.9(9.1)
		$\hat{\Sigma}_u^{bc,thre}$	1	4.6(4.6)	4.1(4.1)	3.3(3.3)	5.7(5.8)	4.9(5.1)	3.4(3.5)	6.7(6.7)	5.6(5.5)	4.4(4.4)
			2	4.7(4.6)	4.1(4.1)	3.3(3.3)	5.9(6.1)	5.0(5.2)	3.4(3.6)	6.5(6.4)	5.4(5.4)	4.3(4.3)
			3	4.8(4.7)	4.2(4.2)	3.3(3.3)	6.3(6.5)	5.2(5.4)	3.6(3.7)	6.3(6.3)	5.3(5.3)	4.2(4.2)
	100	$\hat{\Sigma}_u^{thre}$	1	31.9(32.1)	20.5(20.8)	12.6(12.8)	46.7(48.9)	29.8(31.4)	18.0(19.0)	25.5(25.3)	17.1(17.2)	11.0(10.9)
			2	36.7(36.6)	23.8(23.9)	14.2(14.3)	53.7(56.0)	34.7(36.2)	20.5(21.5)	28.2(28.0)	19.0(19.1)	11.9(11.9)
			3	41.5(41.3)	27.1(27.0)	15.9(16.0)	60.7(63.0)	39.5(41.0)	23.0(24.0)	31.1(30.7)	21.0(21.0)	12.9(12.8)
		$\hat{\Sigma}_u^{bc,thre}$	1	5.4(5.4)	4.5(4.5)	3.7(3.7)	6.4(6.5)	4.9(4.9)	3.7(3.7)	7.5(7.4)	6.3(6.3)	4.9(4.8)
			2	5.4(5.4)	4.5(4.5)	3.7(3.7)	6.6(6.7)	5.0(5.1)	3.7(3.8)	7.3(7.2)	6.1(6.1)	4.8(4.7)
			3	5.6(5.6)	4.6(4.7)	3.7(3.8)	6.9(7.0)	5.2(5.3)	3.8(3.9)	7.2(7.1)	6.0(6.0)	4.7(4.6)
200	$\hat{\Sigma}_u^{thre}$	1	40.8(40.9)	26.5(26.6)	15.5(15.6)	44.7(45.4)	29.0(29.4)	16.8(17.1)	31.0(30.9)	21.0(21.1)	12.8(12.9)	
		2	45.4(45.5)	29.7(29.7)	17.2(17.2)	49.8(50.5)	32.5(32.9)	18.6(18.9)	33.8(33.6)	22.9(22.9)	13.8(13.8)	
		3	50.1(50.1)	32.9(32.8)	18.8(18.8)	55.0(55.6)	35.9(36.3)	20.4(20.6)	36.7(36.5)	24.8(24.8)	14.7(14.8)	
	$\hat{\Sigma}_u^{bc,thre}$	1	7.5(6.2)	5.4(5.1)	4.2(4.1)	8.3(6.4)	5.7(4.8)	3.8(3.6)	8.2(8.2)	6.9(6.9)	5.3(5.3)	
		2	8.1(6.3)	5.7(5.2)	4.2(4.1)	8.9(6.5)	6.1(4.9)	3.9(3.6)	8.1(8.1)	6.7(6.7)	5.2(5.2)	
		3	8.8(6.5)	6.0(5.3)	4.3(4.2)	9.6(6.7)	6.6(5.0)	4.1(3.7)	8.0(8.0)	6.7(6.6)	5.2(5.1)	
Asynchronous data			Model 1			Model 2			Model 3			
$252\sigma_\epsilon^2/\bar{\sigma}^2$	p	Estimators	Δ_K	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$
0.063	50	$\hat{\Sigma}_u^{thre}$	1	6.3(6.3)	5.1(5.1)	3.7(3.7)	9.8(10.1)	7.3(7.6)	4.4(4.7)	9.8(9.7)	7.3(7.2)	5.5(5.5)
			2	6.5(6.4)	5.2(5.2)	3.7(3.7)	9.8(10.3)	7.5(7.8)	4.6(4.8)	9.2(9.2)	7.0(6.9)	5.2(5.2)
			3	6.7(6.7)	5.4(5.3)	3.8(3.8)	10.3(10.7)	7.8(8.2)	4.8(5.1)	8.9(8.8)	6.8(6.7)	5.1(5.1)
		$\hat{\Sigma}_u^{bc,thre}$	1	6.2(6.3)	5.2(5.2)	3.8(3.8)	9.8(10.2)	7.4(7.8)	4.5(4.7)	9.7(9.6)	7.2(7.2)	5.5(5.5)
			2	6.4(6.4)	5.3(5.3)	3.8(3.8)	9.9(10.4)	7.6(8.0)	4.6(4.9)	9.1(9.1)	6.9(6.9)	5.3(5.3)
			3	6.7(6.8)	5.4(5.5)	3.9(3.9)	10.4(11.0)	8.0(8.4)	4.9(5.2)	8.7(8.7)	6.7(6.7)	5.1(5.1)
	100	$\hat{\Sigma}_u^{thre}$	1	7.6(7.6)	5.6(5.6)	4.2(4.2)	11.0(11.3)	7.3(7.4)	4.9(5.0)	10.5(10.4)	8.3(8.2)	5.9(5.9)
			2	7.8(7.8)	5.7(5.7)	4.3(4.3)	11.0(11.3)	7.5(7.7)	5.0(5.1)	10.1(10.0)	7.9(7.8)	5.7(5.7)
			3	8.1(8.1)	5.9(5.9)	4.4(4.4)	11.3(11.6)	7.8(8.0)	5.2(5.3)	9.9(9.8)	7.7(7.6)	5.6(5.6)
		$\hat{\Sigma}_u^{bc,thre}$	1	7.7(7.7)	5.7(5.7)	4.3(4.3)	11.2(11.6)	7.5(7.6)	5.1(5.2)	10.4(10.3)	8.2(8.1)	5.9(5.9)
			2	8.0(8.0)	5.8(5.9)	4.4(4.4)	11.3(11.7)	7.7(7.9)	5.2(5.3)	10.0(9.9)	7.9(7.8)	5.8(5.8)
			3	8.3(8.3)	6.1(6.1)	4.6(4.6)	11.8(12.1)	8.1(8.3)	5.4(5.5)	9.8(9.7)	7.7(7.6)	5.7(5.7)
200	$\hat{\Sigma}_u^{thre}$	1	9.1(9.1)	6.6(6.6)	4.8(4.8)	10.6(11.0)	7.2(7.4)	4.7(4.8)	11.3(11.2)	8.8(8.7)	6.3(6.3)	
		2	9.2(9.2)	6.7(6.7)	4.8(4.9)	10.4(10.8)	7.3(7.4)	4.7(4.8)	11.0(11.0)	8.5(8.5)	6.2(6.2)	
		3	9.4(9.4)	6.8(6.8)	4.9(4.9)	10.6(10.9)	7.4(7.5)	4.7(4.8)	10.9(10.8)	8.4(8.3)	6.1(6.1)	
	$\hat{\Sigma}_u^{bc,thre}$	1	9.4(9.4)	6.9(6.9)	5.0(5.0)	11.1(11.6)	7.6(7.7)	5.0(5.1)	11.2(11.2)	8.7(8.7)	6.4(6.4)	
		2	9.7(9.6)	7.0(7.1)	5.1(5.1)	11.1(11.5)	7.7(7.8)	5.0(5.1)	11.0(11.0)	8.5(8.5)	6.3(6.3)	
		3	10.0(10.0)	7.3(7.3)	5.2(5.2)	11.4(11.7)	7.9(8.1)	5.1(5.2)	10.9(10.9)	8.4(8.4)	6.2(6.2)	

(continued on next page)

Table 1 (continued).

0.00252	$\widehat{\Sigma}_u^{\text{thre}}$	1	61.8(49.2)	46.7(36.8)	18.4(14.7)	109(93.2)	81.8(69.6)	31.5(27.0)	48.2(40.4)	36.7(30.3)	16.1(13.3)	
		2	78.9(62.3)	55.5(43.9)	22.1(17.4)	137(118)	97.2(83.5)	38.2(32.4)	57.5(47.8)	41.5(34.5)	18.3(15.1)	
		3	95.8(75.9)	64.3(51.3)	26.0(20.3)	165(143)	113(97.7)	45.0(38.1)	67.4(55.5)	46.4(38.8)	20.5(16.9)	
	$\widehat{\Sigma}_u^{\text{bc,thre}}$	1	6.9(6.7)	5.4(5.3)	3.8(3.8)	15.1(14.8)	8.1(8.1)	4.7(4.8)	12.7(12.2)	8.5(8.3)	5.8(5.7)	
		2	7.2(6.9)	5.4(5.4)	3.8(3.8)	15.8(15.5)	8.3(8.2)	4.8(5.0)	12.9(12.2)	8.4(8.2)	5.6(5.5)	
		3	7.6(7.0)	5.5(5.5)	3.9(3.9)	16.6(16.6)	8.6(8.5)	5.0(5.3)	13.2(12.3)	8.3(8.1)	5.4(5.4)	
	100	$\widehat{\Sigma}_u^{\text{thre}}$	1	91.9(89.5)	46.8(46.9)	24.7(24.8)	133(136)	68.9(72.1)	35.9(37.6)	65.3(64.0)	37.4(37.1)	20.4(20.3)
			2	107(104)	55.7(55.6)	28.5(28.5)	156(160)	82.1(85.4)	41.6(43.3)	73.9(72.1)	42.1(42.3)	22.6(22.4)
			3	121(119)	64.8(64.8)	32.4(32.2)	179(183)	95.3(98.7)	47.3(49.0)	82.8(81.1)	47.0(47.7)	24.9(24.6)
$\widehat{\Sigma}_u^{\text{bc,thre}}$		1	9.0(8.8)	5.8(5.8)	4.3(4.3)	19.8(21.4)	8.4(8.4)	5.2(5.3)	15.5(15.0)	9.7(9.6)	6.3(6.3)	
		2	9.9(9.9)	5.9(5.9)	4.4(4.4)	20.8(22.8)	8.5(8.5)	5.3(5.4)	16.1(15.5)	9.5(9.4)	6.2(6.1)	
		3	11.3(11.3)	6.1(6.1)	4.5(4.5)	22.0(24.2)	8.8(8.9)	5.5(5.6)	16.9(16.1)	9.5(9.3)	6.1(6.0)	
200	$\widehat{\Sigma}_u^{\text{thre}}$	1	119(121)	63.5(64.5)	31.2(31.4)	132(135)	69.8(71.6)	34.0(34.7)	81.1(82.0)	46.7(47.8)	24.5(24.5)	
		2	132(135)	72.6(73.2)	34.9(35.1)	149(152)	79.5(81.3)	38.1(38.8)	89.9(91.1)	51.7(52.7)	26.6(26.7)	
		3	147(150)	81.8(82.0)	38.7(38.8)	167(170)	89.3(91.1)	42.2(42.9)	99.5(101)	57.1(57.7)	28.9(28.9)	
	$\widehat{\Sigma}_u^{\text{bc,thre}}$	1	14.3(14.3)	7.2(6.7)	5.1(4.9)	25.5(27.1)	9.7(9.7)	5.3(5.1)	18.9(18.9)	10.8(10.7)	6.9(6.8)	
		2	15.6(15.7)	7.4(6.8)	5.2(5.0)	27.2(28.9)	10.0(9.9)	5.5(5.2)	20.1(20.2)	10.8(10.7)	6.7(6.7)	
		3	17.1(17.1)	7.8(6.9)	5.4(5.1)	28.8(30.5)	10.4(10.2)	5.8(5.3)	21.4(21.7)	10.9(10.8)	6.7(6.7)	

different values for $252\sigma_e^2/\bar{\sigma}^2$. However, the FPRs of $\widehat{\Sigma}_u^{\text{thre}}$ when $252\sigma_e^2/\bar{\sigma}^2 = 0.00252$ are much higher than those when $252\sigma_e^2/\bar{\sigma}^2 = 0.063$, showing the impact from weaker signal. This again suggests that the bias-correction is very helpful, especially for handling relatively weaker signals.

6.2. Real data analysis

We analyzed a real high-dimensional data set, studying the statistical properties of microstructure noises that contaminate the trading prices (log-prices) of the constituent stocks of S&P 500. Intra-day tick-by-tick trading data on two days, November 4 and 22, 2016, were downloaded from the TAQ database.

Besides the prices themselves, the Global Industry Classification Standard (GICS) codes are available to classify the companies in S&P 500.³ 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). Since there are only 5 companies belonging to Telecommunication Services, we therefore combined the companies belonging to the Information Technology and Telecommunication Services together and denoted them as ‘T’. Our analysis does not assume any information from the GICS classifications; we use it for validating and interpreting the outcomes from applying the proposed method.

Upon applying the proposed methods, we report in Fig. 1 the magnitudes of the elements in the correlation matrices of the microstructure noises estimated from $\widehat{\Sigma}_u^{\text{bc,thre}}$ in (18), respectively, for November 4 and 22, 2016 with tuning parameters suggested in Section 6.1.3. Here the companies are sorted by the categories defined by the GICS codes. The red blocks along the diagonal in Fig. 1 represent the industrial classifications according to the digits 1–2 of GICS codes. Hence, we can examine both the within- and between-block correlations as revealed by our estimator.

We remark with some interesting findings from Fig. 1. Overall, we can see that the estimated correlation matrices are sparse with many components estimated as zero, indicating that our approach achieved the goal of parsimonious covariance estimations that can (i) effectively identify nonzero components, and (ii) support providing meaningful interpretations. More specifically, we see that the correlations differ substantially on these two days; and such difference is related to the level of the Chicago Board Options Exchange Volatility Index (VIX), a popular measure of the stock market’s expectation of volatility implied by S&P 500 index options. On November 4 when the VIX level was higher, and the overall correlation level between different components of the microstructure noise is also found to be higher than that on November 22. Upon examining the within- and between-category correlations with categories defined by the GICS codes, we see that the correlations within each industrial sector are clear, especially for the Energy and Financial sectors. In contrast, the correlations between different industrial sector are much weaker. Meanwhile, we observe that the correlation estimations have no substantial difference between the cases with $\Delta_K = 1$ and 3, an indication that our method is not sensitive for the choice of Δ_K .

These findings suggest us that it is practically meaningful by studying the high-dimensional statistical properties of the noises. For example, the between-day difference in correlations is helpful to understand the changes in the market sentiment under different market conditions. Furthermore, an interesting feature is the pattern found in the within- and

³ 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.

Table 2
The empirical true positive rates ($\times 100$) and false positive rates ($\times 100$) with synchronous data when jumps exist and do not exist (in parentheses) based on 1000 repetitions.

$252\sigma_e^2/\bar{\sigma}^2$	p	Estimators	Δ_K	Model 1						Model 2					
				TPR			FPR			TPR			FPR		
				$\Delta = 3$	$\Delta = 2$	$\Delta = 1$	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$	$\Delta = 3$	$\Delta = 2$	$\Delta = 1$
0.063	50	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	0.5(0.5)	0.9(0.9)	0.6(0.5)	100(100)	100(100)	100(100)	0.5(0.5)	0.9(0.9)	0.5(0.5)
			2	100(100)	100(100)	100(100)	0.8(0.8)	1.2(1.2)	0.8(0.8)	100(100)	100(100)	100(100)	0.8(0.8)	1.2(1.2)	0.8(0.8)
			3	100(100)	100(100)	100(100)	1.2(1.2)	1.6(1.6)	1.2(1.2)	100(100)	100(100)	100(100)	1.2(1.2)	1.6(1.7)	1.2(1.2)
		$\hat{\Sigma}_u^{bc.thre}$	1	100(100)	100(100)	100(100)	0.5(0.6)	1.0(1.0)	0.6(0.6)	100(100)	100(100)	100(100)	0.5(0.6)	1.0(1.0)	0.6(0.6)
			2	100(100)	100(100)	100(100)	0.8(0.9)	1.4(1.3)	0.9(0.9)	100(100)	100(100)	100(100)	0.8(0.9)	1.3(1.4)	0.9(0.9)
			3	100(100)	100(100)	100(100)	1.3(1.3)	1.8(1.8)	1.4(1.4)	100(100)	100(100)	100(100)	1.3(1.4)	1.8(1.9)	1.4(1.4)
	100	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	0.6(0.6)	0.5(0.4)	0.6(0.5)	100(100)	100(100)	100(100)	0.6(0.6)	0.5(0.5)	0.6(0.6)
			2	100(100)	100(100)	100(100)	0.8(0.8)	0.6(0.6)	0.7(0.7)	100(100)	100(100)	100(100)	0.8(0.8)	0.7(0.6)	0.7(0.7)
			3	100(100)	100(100)	100(100)	1.0(1.0)	0.9(0.9)	1.0(1.0)	100(100)	100(100)	100(100)	1.0(1.0)	0.9(0.9)	1.0(1.0)
		$\hat{\Sigma}_u^{bc.thre}$	1	100(100)	100(100)	100(100)	0.7(0.7)	0.5(0.5)	0.7(0.6)	100(100)	100(100)	100(100)	0.7(0.7)	0.5(0.5)	0.7(0.7)
			2	100(100)	100(100)	100(100)	0.9(0.9)	0.7(0.7)	0.9(0.9)	100(100)	100(100)	100(100)	0.9(0.9)	0.7(0.7)	0.9(0.9)
			3	100(100)	100(100)	100(100)	1.2(1.2)	1.1(1.1)	1.2(1.2)	100(100)	100(100)	100(100)	1.2(1.2)	1.1(1.1)	1.2(1.2)
200	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	0.5(0.5)	0.5(0.5)	0.5(0.5)	100(100)	100(100)	100(100)	0.5(0.5)	0.5(0.5)	0.5(0.5)	
		2	100(100)	100(100)	100(100)	0.6(0.6)	0.6(0.6)	0.6(0.6)	100(100)	100(100)	100(100)	0.6(0.6)	0.6(0.6)	0.6(0.6)	
		3	100(100)	100(100)	100(100)	0.7(0.7)	0.7(0.7)	0.7(0.7)	100(100)	100(100)	100(100)	0.7(0.7)	0.7(0.7)	0.7(0.7)	
	$\hat{\Sigma}_u^{bc.thre}$	1	100(100)	100(100)	100(100)	0.6(0.6)	0.5(0.5)	0.6(0.6)	100(100)	100(100)	100(100)	0.6(0.6)	0.6(0.5)	0.6(0.6)	
		2	100(100)	100(100)	100(100)	0.7(0.8)	0.7(0.7)	0.8(0.8)	100(100)	100(100)	100(100)	0.8(0.8)	0.7(0.7)	0.8(0.8)	
		3	100(100)	100(100)	100(100)	0.9(0.9)	0.8(0.8)	0.9(0.9)	100(100)	100(100)	100(100)	0.9(0.9)	0.9(0.8)	0.9(0.9)	
0.00252	50	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	5.6(4.4)	7.6(6.2)	4.0(3.0)	100(100)	100(100)	100(100)	11.0(9.8)	13.4(12.1)	8.3(7.0)
			2	100(100)	100(100)	100(100)	8.2(6.8)	10.0(8.6)	5.8(4.5)	100(100)	100(100)	100(100)	14.2(13.1)	16.1(15.0)	10.9(9.5)
			3	100(100)	100(100)	100(100)	10.9(9.7)	12.5(11.1)	8.0(6.6)	100(100)	100(100)	100(100)	17.2(16.4)	18.7(17.8)	13.6(12.4)
		$\hat{\Sigma}_u^{bc.thre}$	1	100(100)	100(100)	100(100)	0.5(0.5)	0.8(0.8)	0.5(0.5)	100(100)	100(100)	100(100)	0.5(0.5)	0.8(0.8)	0.5(0.5)
			2	100(100)	100(100)	100(100)	0.7(0.7)	1.1(1.1)	0.8(0.8)	100(100)	100(100)	100(100)	0.7(0.7)	1.0(1.1)	0.7(0.8)
			3	100(100)	100(100)	100(100)	1.1(1.1)	1.5(1.5)	1.2(1.2)	100(100)	100(100)	100(100)	1.1(1.1)	1.5(1.5)	1.1(1.2)
	100	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	4.1(4.0)	3.0(2.8)	2.8(2.7)	100(100)	100(100)	100(100)	7.5(7.3)	6.0(5.8)	5.6(5.3)
			2	100(100)	100(100)	100(100)	5.4(5.2)	4.2(4.0)	3.8(3.6)	100(99.9)	100(100)	100(100)	8.9(8.8)	7.5(7.2)	6.9(6.6)
			3	100(100)	100(100)	100(100)	6.7(6.6)	5.5(5.3)	4.9(4.7)	99.9(99.8)	100(100)	100(100)	10.3(10.2)	9.0(8.7)	8.2(8.0)
		$\hat{\Sigma}_u^{bc.thre}$	1	100(100)	100(100)	100(100)	0.5(0.5)	0.4(0.4)	0.5(0.5)	100(100)	100(100)	100(100)	0.5(0.5)	0.4(0.4)	0.5(0.5)
			2	100(100)	100(100)	100(100)	0.7(0.7)	0.6(0.6)	0.7(0.7)	100(100)	100(100)	100(100)	0.7(0.7)	0.6(0.6)	0.7(0.7)
			3	100(100)	100(100)	100(100)	0.9(0.9)	0.8(0.8)	1.0(1.0)	100(100)	100(100)	100(100)	0.9(0.9)	0.8(0.8)	1.0(1.0)
200	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	2.8(2.8)	2.2(2.2)	2.0(2.0)	99.8(99.8)	100(100)	100(100)	4.6(4.7)	3.9(4.0)	3.6(3.6)	
		2	100(100)	100(100)	100(100)	3.3(3.4)	2.7(2.8)	2.4(2.5)	99.8(99.7)	100(100)	100(100)	5.2(5.3)	4.6(4.6)	4.1(4.2)	
		3	100(100)	100(100)	100(100)	3.9(3.9)	3.3(3.4)	2.9(3.0)	99.6(99.6)	100(99.9)	100(100)	5.8(5.8)	5.2(5.2)	4.7(4.7)	
	$\hat{\Sigma}_u^{bc.thre}$	1	100(100)	100(100)	100(100)	0.5(0.5)	0.4(0.4)	0.5(0.5)	100(100)	100(100)	100(100)	0.5(0.5)	0.4(0.4)	0.5(0.5)	
		2	100(100)	100(100)	100(100)	0.6(0.6)	0.5(0.5)	0.6(0.6)	100(100)	100(100)	100(100)	0.6(0.6)	0.5(0.5)	0.6(0.6)	
		3	100(100)	100(100)	100(100)	0.7(0.7)	0.7(0.7)	0.8(0.8)	100(100)	100(100)	100(100)	0.7(0.7)	0.7(0.7)	0.8(0.8)	

Table 3
The empirical true positive rates ($\times 100$) and false positive rates ($\times 100$) in different settings with asynchronous data when jumps exist and do not exist (in parentheses) based on 1000 repetitions.

252 $\sigma_e^2/\bar{\sigma}^2$	p	Estimators	Δ_K	Model 1						Model 2					
				TPR			FPR			TPR			FPR		
				$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$
0.063	50	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	0.6(0.6)	0.9(0.9)	0.4(0.4)	99.3(98.9)	100(100)	100(100)	0.6(0.6)	0.9(0.9)	0.5(0.4)
			2	100(100)	100(100)	100(100)	0.9(0.9)	1.2(1.2)	0.7(0.7)	99.7(99.5)	100(100)	100(100)	0.9(0.9)	1.3(1.2)	0.7(0.7)
			3	100(100)	100(100)	100(100)	1.3(1.3)	1.6(1.6)	1.1(1.1)	99.8(99.7)	100(100)	100(100)	1.4(1.3)	1.7(1.7)	1.1(1.1)
		$\hat{\Sigma}_u^{bc,thre}$	1	100(100)	100(100)	100(100)	0.6(0.6)	1.0(1.0)	0.5(0.5)	99.3(98.9)	100(100)	100(100)	0.6(0.6)	1.0(1.0)	0.5(0.5)
			2	100(100)	100(100)	100(100)	1.0(1.0)	1.4(1.4)	0.8(0.8)	99.7(99.4)	100(100)	100(100)	1.0(1.0)	1.4(1.4)	0.8(0.8)
			3	100(100)	100(100)	100(100)	1.5(1.5)	1.9(1.9)	1.3(1.3)	99.8(99.7)	100(100)	100(100)	1.5(1.5)	1.9(1.9)	1.3(1.3)
	100	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	0.6(0.6)	0.4(0.4)	0.5(0.5)	99.4(99.1)	100(100)	100(100)	0.6(0.6)	0.4(0.4)	0.5(0.5)
			2	100(100)	100(100)	100(100)	0.8(0.8)	0.6(0.6)	0.7(0.7)	99.7(99.5)	100(100)	100(100)	0.9(0.9)	0.7(0.7)	0.7(0.7)
			3	100(100)	100(100)	100(100)	1.1(1.1)	0.9(0.9)	1.0(1.0)	99.8(99.7)	100(100)	100(100)	1.1(1.1)	0.9(0.9)	1.0(1.0)
		$\hat{\Sigma}_u^{bc,thre}$	1	100(100)	100(100)	100(100)	0.7(0.7)	0.5(0.5)	0.6(0.6)	99.4(99.0)	100(100)	100(100)	0.7(0.7)	0.5(0.5)	0.6(0.6)
			2	100(100)	100(100)	100(100)	1.0(1.0)	0.8(0.8)	0.9(0.9)	99.7(99.4)	100(100)	100(100)	1.0(1.0)	0.8(0.8)	0.9(0.8)
			3	100(100)	100(100)	100(100)	1.3(1.3)	1.1(1.1)	1.2(1.2)	99.8(99.6)	100(100)	100(100)	1.3(1.3)	1.1(1.1)	1.2(1.2)
	200	$\hat{\Sigma}_u^{thre}$	1	100(100)	100(100)	100(100)	0.5(0.5)	0.5(0.5)	0.5(0.5)	99.4(98.8)	100(100)	100(100)	0.5(0.5)	0.5(0.5)	0.5(0.5)
			2	100(100)	100(100)	100(100)	0.6(0.6)	0.6(0.6)	0.6(0.6)	99.6(99.2)	100(100)	100(100)	0.6(0.6)	0.6(0.6)	0.6(0.6)
			3	100(100)	100(100)	100(100)	0.7(0.7)	0.7(0.7)	0.7(0.7)	99.7(99.4)	100(100)	100(100)	0.7(0.7)	0.7(0.7)	0.7(0.7)
		$\hat{\Sigma}_u^{bc,thre}$	1	100(100)	100(100)	100(100)	0.6(0.6)	0.6(0.6)	0.6(0.6)	99.3(98.6)	100(100)	100(100)	0.6(0.6)	0.6(0.6)	0.6(0.6)
			2	100(100)	100(100)	100(100)	0.8(0.8)	0.7(0.7)	0.8(0.8)	99.6(99.1)	100(100)	100(100)	0.8(0.8)	0.7(0.7)	0.8(0.8)
			3	100(100)	100(100)	100(100)	1.0(1.0)	0.9(0.9)	0.9(0.9)	99.7(99.3)	100(100)	100(100)	1.0(1.0)	0.9(0.9)	0.9(0.9)
0.00252	50	$\hat{\Sigma}_u^{thre}$	1	99.8(100)	100(100)	100(100)	6.6(5.6)	8.9(7.9)	4.7(3.6)	88.4(89.8)	97.0(97.9)	100(100)	12.7(12.0)	15.3(14.6)	9.8(8.5)
			2	98.4(99.6)	100(100)	100(100)	9.3(8.4)	11.6(10.7)	7.1(5.7)	87.8(89.9)	96.6(97.4)	100(100)	15.8(15.4)	18.1(17.6)	12.9(11.7)
			3	95.5(98.7)	100(100)	100(100)	11.8(11.2)	14.1(13.4)	9.8(8.4)	87.1(89.7)	96.3(97.0)	100(100)	18.4(18.3)	20.5(20.4)	15.9(14.9)
		$\hat{\Sigma}_u^{bc,thre}$	1	100(100)	100(100)	100(100)	0.2(0.2)	0.5(0.5)	0.4(0.4)	91.8(92.9)	99.8(99.8)	100(100)	0.2(0.2)	0.4(0.5)	0.4(0.4)
			2	100(100)	100(100)	100(100)	0.3(0.3)	0.6(0.6)	0.6(0.6)	91.2(92.7)	99.8(99.8)	100(100)	0.3(0.3)	0.6(0.6)	0.6(0.6)
			3	99.9(100)	100(100)	100(100)	0.3(0.4)	0.7(0.8)	0.9(0.9)	90.0(92.3)	99.7(99.8)	100(100)	0.3(0.4)	0.7(0.8)	0.9(0.9)
	100	$\hat{\Sigma}_u^{thre}$	1	96.2(97.4)	100(100)	100(100)	4.5(4.4)	3.6(3.5)	3.6(3.4)	86.1(80.9)	98.1(97.4)	100(100)	8.2(8.1)	7.0(6.9)	6.8(6.6)
			2	93.4(94.8)	100(100)	100(100)	5.6(5.6)	5.0(4.9)	4.8(4.6)	85.4(80.2)	97.9(97.0)	100(100)	9.3(9.3)	8.6(8.4)	8.2(8.1)
			3	89.9(91.8)	100(99.8)	100(100)	6.6(6.6)	6.3(6.2)	6.1(6.0)	84.3(79.3)	97.7(96.6)	100(100)	10.3(10.3)	10.0(9.9)	9.7(9.5)
		$\hat{\Sigma}_u^{bc,thre}$	1	99.9(99.9)	100(100)	100(100)	0.2(0.2)	0.2(0.2)	0.4(0.4)	87.5(84.5)	99.8(99.6)	100(100)	0.2(0.2)	0.2(0.2)	0.4(0.4)
			2	99.8(99.8)	100(100)	100(100)	0.2(0.2)	0.3(0.3)	0.6(0.6)	86.5(83.4)	99.8(99.7)	100(100)	0.2(0.2)	0.3(0.3)	0.6(0.6)
			3	99.5(99.5)	100(100)	100(100)	0.2(0.2)	0.4(0.4)	0.8(0.8)	85.1(82.0)	99.8(99.7)	100(100)	0.2(0.2)	0.4(0.4)	0.8(0.8)
	200	$\hat{\Sigma}_u^{thre}$	1	89.3(89.6)	100(99.9)	100(100)	2.7(2.8)	2.6(2.6)	2.5(2.6)	76.1(70.5)	98.0(97.0)	100(99.9)	4.7(4.7)	4.4(4.5)	4.3(4.4)
			2	85.6(86.0)	99.9(99.8)	100(100)	3.1(3.2)	3.1(3.2)	3.0(3.1)	73.9(68.4)	97.9(96.9)	100(99.9)	5.0(5.1)	5.0(5.1)	4.9(5.0)
			3	82.4(82.7)	99.5(99.7)	100(100)	3.5(3.5)	3.6(3.7)	3.6(3.7)	71.9(66.6)	97.7(96.7)	99.9(99.8)	5.4(5.5)	5.6(5.6)	5.5(5.6)
		$\hat{\Sigma}_u^{bc,thre}$	1	99.1(99.1)	100(100)	100(100)	0.1(0.1)	0.2(0.2)	0.4(0.4)	76.1(70.9)	99.2(98.6)	100(100)	0.1(0.1)	0.2(0.2)	0.4(0.4)
			2	98.4(98.3)	100(100)	100(100)	0.1(0.1)	0.2(0.2)	0.5(0.5)	73.7(68.5)	99.2(98.5)	100(100)	0.1(0.1)	0.2(0.2)	0.5(0.5)
			3	97.4(97.3)	100(100)	100(100)	0.1(0.1)	0.3(0.3)	0.6(0.6)	71.4(66.3)	99.1(98.4)	100(100)	0.1(0.1)	0.3(0.3)	0.6(0.6)

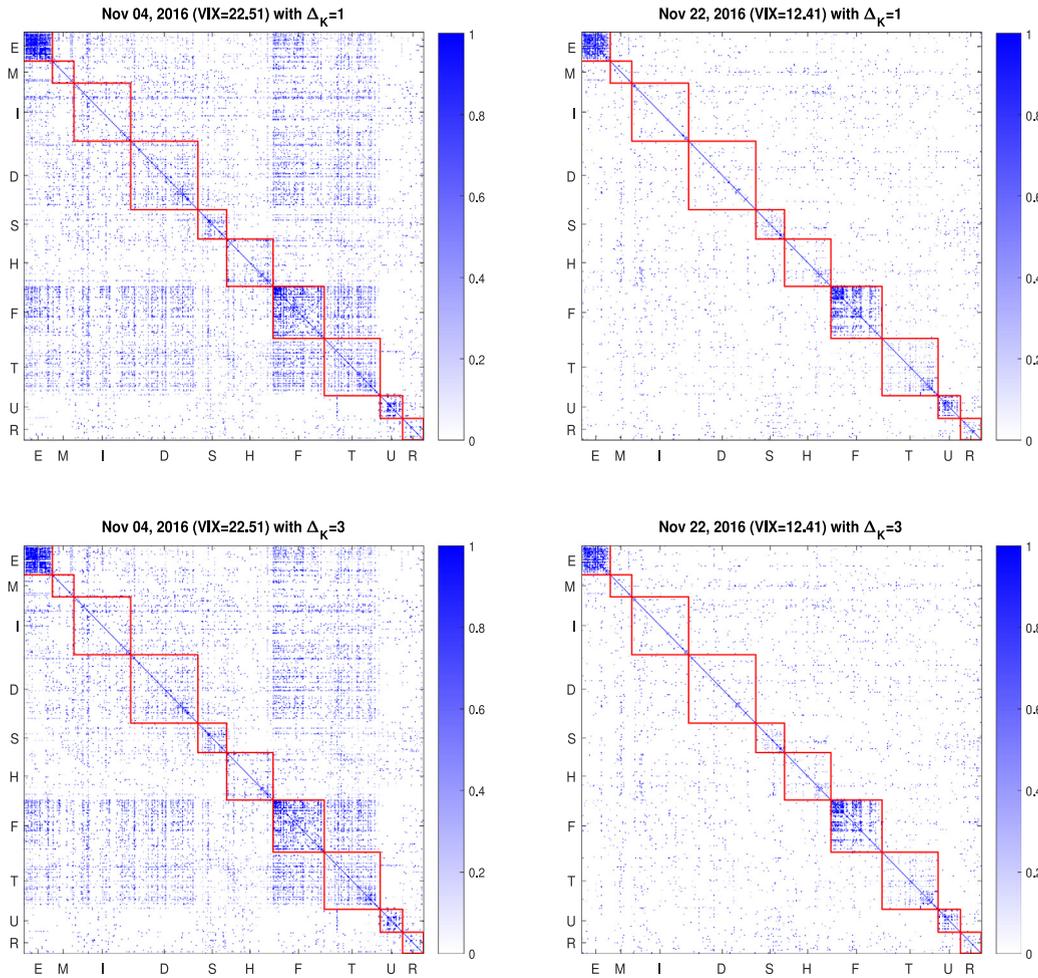


Fig. 1. Magnitudes of the elements in the estimated correlation matrix of the microstructure noises on November 4 and 22, 2016. Note: The graph describes the magnitudes of the elements in the estimated correlation matrices of the microstructure noises. Different colors denote different values of the pairwise correlations. The red squares along the diagonal denote the sectors—Energy (E), Materials (M), Industrials (I), Consumer Discretionary (D), Consumer Staples (S), Health Care (H), Financial (F), Information Technology and Telecommunication Services (T), Utilities (U), and Real Estate (R).

between-industrial sector correlations, which is seen interrelated to the market conditions. Broad questions include how the correlations between noises vary associated with the prices and/or the volatility of different assets, how the sparse covariance matrix of the noises can help in solving practical problems, and so on. Supported by our new methods, we expect more future investigations along this line.

7. 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 (1), the method developed in this paper could be applied to other types of process \mathbf{X}_t , such as the smooth ones typically encountered in the functional data literature. The key property that makes our method work 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 (1). On the other hand, Assumption 2 can be replaced by a weaker assumption that $\min_{i,j \in [p]} n_{i,j} \rightarrow \infty$ and $\max_{i,j \in [p]} \max_{k \in [n_{i,j}-1]} \Delta t_{i,j,k} \rightarrow 0$ as $n \rightarrow \infty$. If we write $n_{\min} = \min_{i,j \in [p]} n_{i,j}$ and assume $\max_{i,j \in [p]} \max_{k \in [n_{i,j}-1]} \Delta t_{i,j,k} \asymp n_{\min}^{-\epsilon}$ for some $\epsilon \in (0, 1]$, with suitable selection of K , Theorems 1 and 3 in Section 3 still hold with replacing n_* by $n_{\min}^{g(\epsilon)}$, where $g(\epsilon) \in (0, 1]$ is a function of ϵ . More specifically, if $\epsilon = 1$, then $g(1) = 1$. However, whether such rates are minimax optimal under the associated losses or not is unclear.

In the analysis of this study, we focus on noises with homoscedastic covariance matrix, so that our target is $\text{Cov}(\mathbf{U}_{t_k}) \equiv \Sigma_u$ for each $k \in [n]$. As an interesting problem, our framework can be extensively developed to handle time-varying heteroscedastic noise where the target covariance is time-dependent $\text{Cov}(\mathbf{U}_{t_s})$ at a given time point t_s instead. Denote by \circ the Hadamard product. Assume $\mathbf{U}_{t_k} = \boldsymbol{\gamma}_{t_k} \circ \boldsymbol{\chi}_{t_k}$ with $\boldsymbol{\gamma}_{t_k} = (\gamma_{1,t_k}, \dots, \gamma_{p,t_k})^\top$ and $\boldsymbol{\chi}_{t_k} = (\chi_{1,t_k}, \dots, \chi_{p,t_k})^\top$, where $\gamma_{1,t}, \dots, \gamma_{p,t}$ are p nonnegative continuous-time processes, and each sequence $\{\chi_{j,t_k}\}_{k \geq 1}$ is α -mixing. Without loss of generality, we assume $\mathbb{E}(\chi_{j,t_k}) \equiv 0$ and $\mathbb{E}(\chi_{j,t_k}^2) = 1$ for each $j \in [p]$ and $k \in [n]$. Recall $\mathcal{G}_i \cap \mathcal{G}_j = \{t_{i,j,1}, \dots, t_{i,j,n_{i,j}}\}$ with $t_{i,j,1} < \dots < t_{i,j,n_{i,j}}$ for any $i, j \in [p]$. We outline a framework as follows.

Step 1. Given some $\tilde{\xi} = o(1)$, define $\mathcal{G}_i^{(s)} = \{t_{i,\ell} \in \mathcal{G}_i : |t_{i,\ell} - t_s| \leq \tilde{\xi}\}$ and $n_{i,j}^{(s)} = |\mathcal{G}_i^{(s)} \cap \mathcal{G}_j^{(s)}|$. For each $t_{i,j,k} \in \mathcal{G}_i^{(s)} \cap \mathcal{G}_j^{(s)}$, write $S_{i,j,k}^{(s)} = \{t_{i,j,\ell} \in \mathcal{G}_i^{(s)} \cap \mathcal{G}_j^{(s)} : K \leq |\ell - k| \leq K + \Delta_K\}$ and $N_{i,j,k}^{(s)} = |S_{i,j,k}^{(s)}|$. We then define $\hat{\sigma}_{u,i,j}^{(s)}$ in the same manner as $\hat{\sigma}_{u,i,j}$ in (5) with replacing $(n_{i,j}, N_{i,j,k}, S_{i,j,k})$ by $(n_{i,j}^{(s)}, N_{i,j,k}^{(s)}, S_{i,j,k}^{(s)})$. Under the independence between the process $\boldsymbol{\gamma}_t$ and the sequence $\{\boldsymbol{\chi}_{t_k}\}_{k \geq 1}$, following our current technical arguments, we have $\hat{\sigma}_{u,i,j}^{(s)} \rightarrow \gamma_{i,t_s} \gamma_{j,t_s} \mathbb{E}(\chi_{i,t_s} \chi_{j,t_s})$ in probability with suitable selection of $\tilde{\xi}$.

Step 2. Given some integers $\tilde{K} \geq 1$ and $h \geq 1$, if $\mathbb{E}(\gamma_{i,t_k} \gamma_{j,t_k})$ and $\mathbb{E}(\chi_{i,t_k} \chi_{j,t_k})$ are both slowly varying with k , we know $(\tilde{N}_{i,j}^{(s)})^{-1} \sum_{b=-\tilde{K}}^{\tilde{K}} \hat{\sigma}_{u,i,j}^{(s+bh)}$ with $\tilde{N}_{i,j}^{(s)} = \{s + bh \in [n] : |b| \leq \tilde{K}\}$ will provide a consistent estimator for $\text{Cov}(U_{i,t_s}, U_{j,t_s})$ under some regularity conditions with suitable selections of \tilde{K} and h .

Clearly, this problem differs substantially from our current investigation. The technical analysis of such estimator would require an extensive framework including additional assumptions on $\{\boldsymbol{\gamma}_{t_k}\}_{k \geq 1}$ and $\{\boldsymbol{\chi}_{t_k}\}_{k \geq 1}$ which are beyond the scope of this study. We plan to carefully investigate this problem in a future project.

8. Proofs

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

8.1. Proof of Theorem 1

For any $k \in [n_{i,j}]$, let $S_{i,j,k} = \{t_{i,j,\ell} : K \leq |\ell - k| \leq K + \Delta_K\}$. For any $i, j \in [p]$, we have that

$$\begin{aligned} \hat{\sigma}_{u,i,j} - \sigma_{u,i,j} &= \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (U_{i,t_{i,j,\ell}} - U_{i,t_{i,j,k}})(U_{j,t_{i,j,\ell}} - U_{j,t_{i,j,k}})}_{\text{I}(i,j)} - \sigma_{u,i,j} \\ &+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}})(X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}})}_{\text{II}(i,j)} \\ &+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}})(U_{j,t_{i,j,\ell}} - U_{j,t_{i,j,k}})}_{\text{III}(i,j)} \\ &+ \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (U_{i,t_{i,j,\ell}} - U_{i,t_{i,j,k}})(X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}})}_{\text{IV}(i,j)}. \end{aligned}$$

Define $\xi = \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} |t_{i,j,\ell} - t_{i,j,k}|$. To prove Theorem 1, we need the following three propositions whose proofs are given in Sections 8.2–8.4, respectively.

Proposition 1. Under Assumptions 1–3, we have that

$$\max_{i,j \in [p]} \mathbb{P}\{|I(i,j)| > v\} \lesssim \{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp\{-C(n_* L_n^{-1} \rho^{-1} v)^\rho / (\varphi + 1)\} + v^{-1} \exp(-Cn_* K^{-1} \rho^{-1} v)$$

for any $v \gg \exp(-CL_n^{-\varphi} K^\varphi)$ and $\rho \geq 1$. Furthermore, it holds that $\max_{i,j \in [p]} \mathbb{E}\{|I(i,j)|^m\} \lesssim 1$ for any fixed positive integer m .

Remark 5. As shown in Section 8.2, the upper bound stated in Proposition 1 holds for any $v \gg \delta := \max_{i,j \in [p]} |n_{i,j}^{-1} \sum_{k=1}^{n_{i,j}} \mathbb{E}(N_{i,j,k}^{-1} U_{i,t_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}})|$. Since $\delta \lesssim \exp(-CL_n^{-\varphi} K^\varphi)$, by Davydov's inequality, we need the restriction $v \gg \exp(-CL_n^{-\varphi} K^\varphi)$ in general settings.

(i) If $\{\mathbf{U}_{t_k}\}$ is an independent sequence, we can select $L_n = 1/2$ and $\varphi = \infty$ in [Assumption 1](#). Then we have $\delta = 0$ in this setting, and the upper bound stated in [Proposition 1](#) can be refined as $\{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp(-Cn_*K^{-1} \rho^{-1} v)$ which holds for any $v > 0$ and $\rho \geq 1$.

(ii) If $\{\mathbf{U}_{t_k}\}$ is an L_n -dependent sequence, we can select $\varphi = \infty$ in [Assumption 1](#). If $K > L_n$, we have $\delta = 0$ in this setting, and the upper bound stated in [Proposition 1](#) can be refined as $\{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp(-Cn_*K^{-1} \rho^{-1} v)$ which holds for any $v > 0$ and $\rho \geq 1$.

Proposition 2. Under [Assumptions 2, 4 and 5](#), we have that

$$\max_{i,j \in [p]} \mathbb{P}\{|\text{II}(i, j)| > v\} \lesssim \exp(-Cn_*K^{-1}v^2) + \exp\{-C(n_*K^{-1})^{\gamma/(\gamma+4)}\}$$

for any $v = o\{(n_*^{-1}K)^{2/(\gamma+4)}\}$, where γ is specified in [Assumption 5](#). Furthermore, it holds that $\max_{i,j \in [p]} \mathbb{E}\{|\text{II}(i, j)|^m\} \lesssim n_*^{-m}K^m$ for any fixed positive integer m .

Proposition 3. Under [Assumptions 1–5](#), if $K(\log n_*)^{1+2/\gamma} = o(n_*)$ and $K \gtrsim L_n$, we have that

$$\max_{i,j \in [p]} \mathbb{P}\{|\text{III}(i, j)| \geq v\} \lesssim \exp(-Cn_*L_n^{-1}v^2) + \exp\{-C(n_*L_n^{-1}v)^{\varphi/(2\varphi+1)}\} + \exp\{-C(n_*K^{-1})^{\gamma/(\gamma+2)}\}$$

for any $v > 0$, where γ is specified in [Assumption 5](#). Furthermore, it holds that $\max_{i,j \in [p]} \mathbb{E}\{|\text{III}(i, j)|^m\} \lesssim n_*^{-m/2}K^{m/2}$ for any fixed positive integer m . The same results also hold for $\text{IV}(i, j)$.

Remark 6. If $\varphi = \infty$, the upper bound stated in [Proposition 3](#) can be refined as $\exp(-Cn_*L_n^{-1}v^2) + \exp\{-C(n_*L_n^{-1}v)^{1/2}\} + \exp\{-C(n_*K^{-1})^{\gamma/(\gamma+2)}\}$.

Write $\aleph = (n_*^{-1}K \log p)^{1/2}$. We first consider the case with $\varphi < \infty$. Notice that $(1 + cx^{-1})^{-x} \geq e^{-c}$ for any $x > 0$ and $c > 0$. By [Propositions 1–3](#), if $K(\log n_*)^{1+2/\gamma} = o(n_*)$ and $K \gtrsim L_n$, we have

$$\begin{aligned} \max_{i,j \in [p]} \mathbb{P}\{|\hat{\sigma}_{u,ij} - \sigma_{u,ij}| > v\} &\lesssim \{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp\{-C(n_*L_n^{-1} \rho^{-1} v)^{\varphi/(\varphi+1)}\} \\ &\quad + v^{-1} \exp(-Cn_*K^{-1} \rho^{-1} v) + \exp\{-C(n_*K^{-1})^{\gamma/(\gamma+4)}\} + \exp\{-C(n_*L_n^{-1}v)^{\varphi/(2\varphi+1)}\} \end{aligned} \tag{19}$$

for any $v \gg \exp(-CL_n^{-\varphi}K^\varphi)$, $v = o\{(n_*^{-1}K)^{2/(\gamma+4)}\}$ and $\rho \geq 1$. Since $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$ and $\log p = o\{(n_*K^{-1})^{\gamma/(\gamma+4)}\}$, then $\exp(-CL_n^{-\varphi}K^\varphi) = o(\aleph)$ and $\aleph = o\{(n_*^{-1}K)^{2/(\gamma+4)}\}$. Given a sufficiently large constant $\alpha > 0$, it holds that

$$\begin{aligned} \mathbb{E}\{|\widehat{\Sigma}_u - \Sigma_u|_\infty\} &\leq \mathbb{E}\left\{\max_{i,j \in [p]} |\hat{\sigma}_{u,ij} - \sigma_{u,ij}| I(|\hat{\sigma}_{u,ij} - \sigma_{u,ij}| \leq \alpha \aleph)\right\} + \mathbb{E}\left\{\max_{i,j \in [p]} |\hat{\sigma}_{u,ij} - \sigma_{u,ij}| I(|\hat{\sigma}_{u,ij} - \sigma_{u,ij}| > \alpha \aleph)\right\} \\ &=: A_1 + A_2. \end{aligned}$$

It is easy to see that $A_1 \leq \alpha \aleph$. By Cauchy–Schwarz inequality, we have

$$\begin{aligned} A_2 &\leq \sum_{i,j=1}^p \mathbb{E}\{|\hat{\sigma}_{u,ij} - \sigma_{u,ij}| I(|\hat{\sigma}_{u,ij} - \sigma_{u,ij}| > \alpha \aleph)\} \\ &\leq p^2 \max_{i,j \in [p]} \{ \mathbb{E}\{(|\hat{\sigma}_{u,ij} - \sigma_{u,ij}|^2)\} \}^{1/2} \cdot \max_{i,j \in [p]} \{ \mathbb{P}\{|\hat{\sigma}_{u,ij} - \sigma_{u,ij}| > \alpha \aleph\} \}^{1/2}. \end{aligned}$$

Let $\rho \asymp \log p \geq 1$. Since $K \log p = o(n_*)$, then it follows from [\(19\)](#) that

$$\begin{aligned} \max_{i,j \in [p]} \mathbb{P}\{|\hat{\sigma}_{u,ij} - \sigma_{u,ij}| > \alpha \aleph\} &\lesssim p^{-2w} + \exp\{-C(n_*KL_n^{-2} \log p)^{\varphi/(4\varphi+2)}\} + \exp[-C\{n_*K^{-1}(\log p)^{-1}\}^{1/2}] \\ &\quad + \exp\{-C(n_*K^{-1})^{\gamma/(\gamma+4)}\} + \exp[-C\{n_*KL_n^{-2}(\log p)^{-1}\}^{\varphi/(2\varphi+2)}] \end{aligned}$$

with some sufficiently large $w > 0$, where $w \rightarrow \infty$ as $\alpha \rightarrow \infty$. Due to $\max_{i,j \in [p]} \mathbb{E}\{|\hat{\sigma}_{u,ij} - \sigma_{u,ij}|^2\} \lesssim 1$, if $\log p = o[\min\{(n_*L_n^{-2}K)^\varphi/(3\varphi+2), (n_*K^{-1})^\chi\}]$ with $\chi = \min\{\gamma/(\gamma + 4), 1/3\}$, then

$$\begin{aligned} A_2 &\lesssim p^{2-w} + p^2 \exp\{-C(n_*KL_n^{-2} \log p)^{\varphi/(4\varphi+2)}\} + p^2 \exp[-C\{n_*K^{-1}(\log p)^{-1}\}^{1/2}] \\ &\quad + p^2 \exp\{-C(n_*K^{-1})^{\gamma/(\gamma+4)}\} + p^2 \exp[-C\{n_*KL_n^{-2}(\log p)^{-1}\}^{\varphi/(2\varphi+2)}] \\ &\lesssim p^{2-w} + \exp\{-C(n_*KL_n^{-2} \log p)^{\varphi/(4\varphi+2)}\} + \exp[-C\{n_*K^{-1}(\log p)^{-1}\}^{1/2}] \\ &\quad + \exp\{-C(n_*K^{-1})^{\gamma/(\gamma+4)}\} + \exp[-C\{n_*KL_n^{-2}(\log p)^{-1}\}^{\varphi/(2\varphi+2)}] \\ &= o\{(n_*^{-1}K \log p)^4\}. \end{aligned}$$

Hence, $\sup_{\mathcal{P}_1} \mathbb{E}\{|\widehat{\Sigma}_u - \Sigma_u|_\infty\} \lesssim (n_*^{-1}K \log p)^{1/2}$ provided that $K(\log n_*)^{1+2/\gamma} = o(n_*)$, $K \gtrsim L_n$, $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$ and $\log p = o[\min\{(n_*L_n^{-2}K)^\varphi/(3\varphi+2), (n_*K^{-1})^\chi\}]$.

Now we consider the case with $\varphi = \infty$. As we discussed in Remark 5(i), if $\{\mathbf{U}_{t_k}\}$ is an independent sequence, we can select $L_n = 1/2$. Due to $K \geq 1$, we have $K > L_n$ in this case. Without loss of generality, we can always assume $K > L_n$ when $\varphi = \infty$. Based on Remark 5, it holds that $\{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp(-Cn_*K^{-1} \rho^{-1} v)$ for any $v > 0$ and $\rho \geq 1$ under either of the scenarios: (i) $\{\mathbf{U}_{t_k}\}$ is an independent sequence, and (ii) $\{\mathbf{U}_{t_k}\}$ is an L_n -dependent sequence. Repeating the arguments for $\varphi < \infty$, we have $\sup_{\mathcal{P}_1} \mathbb{E}(|\widehat{\Sigma}_u - \Sigma_u|_\infty) \lesssim (n_*^{-1} K \log p)^{1/2}$ provided that $K(\log n_*)^{1+2/\gamma} = o(n_*)$, $K > L_n$ and $\log p = o\{(n_*K^{-1})^x\}$. We complete the proof of Theorem 1. \square

8.2. Proof of Proposition 1

To prove Proposition 1, we need the following lemma whose proof is given in Section 8.16.

Lemma 1. Let $\{z_t\}_{t=1}^{\tilde{n}}$ be an α -mixing sequence of real-valued and centered random variables with α -mixing coefficients $\{\alpha(k)\}_{k \geq 1}$. Assume there exist some universal constants $a_1 > 1$, $a_2, b_1, b_2 > 0$, $r \in (0, 2]$ and $\varphi > 0$ such that (i) $\max_{t \in [\tilde{n}]} \mathbb{P}(|z_t| > u) \leq b_1 \exp(-b_2 u^r)$ for any $u > 0$, (ii) $\alpha(k) \leq a_1 \exp(-a_2 \tilde{L}_{\tilde{n}}^{-\varphi} |k - m|_+^\varphi)$ for any integer $k \geq 1$, where $\tilde{L}_{\tilde{n}} > 0$ and $m = o(\tilde{n}) \geq 0$ may diverge with \tilde{n} . Let $s_{\tilde{n}}^2 = \sum_{t_1, t_2=1}^{\tilde{n}} |\text{Cov}(z_{t_1}, z_{t_2})|$ and $r_* = r\varphi/(r + \varphi)$. It holds that

$$\mathbb{P}\left(\sup_{k \in [\tilde{n}]} \left| \sum_{t=1}^k z_t \right| \geq \tilde{n}x\right) \lesssim (1 + \tilde{n}^2 x^2 \rho^{-1} s_{\tilde{n}}^{-2})^{-\rho/2} + x^{-1} \exp(-C\tilde{n}^{r_*} x^{r_*} \rho^{-r_*} \tilde{L}_{\tilde{n}}^{-r_*}) + x^{-1} \exp(-Cm^{-r} \tilde{n}^r x^r \rho^{-r})$$

for any $x > 0$ and $\rho \geq 1$, where we adopt the convention $\exp(-C0^{-r} x^r) = 0$ for any $x > 0$.

Remark 7. If $\varphi = \infty$, we have $r_* = r$. Then the upper bound in Lemma 1 can be refined as $(1 + \tilde{n}^2 x^2 \rho^{-1} s_{\tilde{n}}^{-2})^{-\rho/2} + x^{-1} \exp(-C\tilde{n}^r x^r \rho^{-r} (m + \tilde{L}_{\tilde{n}})^{-r})$.

Now we begin to prove Proposition 1. Recall that

$$\begin{aligned} I(i, j) &= \underbrace{\frac{1}{n_{i,j}} \sum_{\ell=1}^{n_{i,j}} \left(\frac{1}{2} + \frac{1}{2} \sum_{t_{i,j,k} \in S_{i,j,\ell}} \frac{1}{N_{i,j,k}} \right)}_{I_1(i,j)} (U_{i,t_{i,j,\ell}} U_{j,t_{i,j,\ell}} - \sigma_{u,i,j}) \\ &\quad - \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{U_{i,t_{i,j,k}}}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}}_{I_2(i,j)} - \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{U_{j,t_{i,j,k}}}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{i,t_{i,j,\ell}}}_{I_3(i,j)}. \end{aligned} \tag{20}$$

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

For each $\ell \in [n_{i,j}]$, let $\zeta_{i,j,\ell} = 2^{-1} (1 + \sum_{t_{i,j,k} \in S_{i,j,\ell}} N_{i,j,k}^{-1}) (U_{i,t_{i,j,\ell}} U_{j,t_{i,j,\ell}} - \sigma_{u,i,j})$. Then we have $I_1(i, j) = n_{i,j}^{-1} \sum_{\ell=1}^{n_{i,j}} \zeta_{i,j,\ell}$. Recall that $N_{i,j,k} = |S_{i,j,k}|$ with $S_{i,j,k} = \{t_{i,j,\ell} : K \leq |\ell - k| \leq K + \Delta_K\}$. Since $K < K + \Delta_K = o(n_*)$ and $n_{i,j} \asymp n_* \rightarrow \infty$ as $n \rightarrow \infty$, we then have $2(K + \Delta_K) < n_{i,j}$ for sufficiently large n . Thus for sufficiently large n , it holds that

$$N_{i,j,k} = \begin{cases} \Delta_K + 1, & \text{if } 1 \leq k \leq K, \\ \Delta_K - K + k + 1, & \text{if } K + 1 \leq k \leq K + \Delta_K, \\ 2\Delta_K + 2, & \text{if } K + \Delta_K + 1 \leq k \leq n_{i,j} - K - \Delta_K, \\ n_{i,j} + \Delta_K - K - k + 2, & \text{if } n_{i,j} - K - \Delta_K + 1 \leq k \leq n_{i,j} - K, \\ \Delta_K + 1, & \text{if } n_{i,j} - K + 1 \leq k \leq n_{i,j}, \end{cases} \tag{21}$$

which implies that $\min_{k \in [n_{i,j}]} N_{i,j,k} = \Delta_K + 1$ and $\max_{k \in [n_{i,j}]} N_{i,j,k} = 2\Delta_K + 2$. Therefore, we have that $C^{-1} < \sum_{t_{i,j,k} \in S_{i,j,\ell}} N_{i,j,k}^{-1} < C$ holds uniformly over $\ell \in [n_{i,j}]$ and $i, j \in [p]$. By Lemma 2 of Chang et al. (2013), Assumption 3 yields $\max_{i,j \in [p], \ell \in [n_{i,j}]} \mathbb{P}(|\zeta_{i,j,\ell}| > v) \leq C \exp(-Cv)$ for any $v > 0$, which implies $\max_{i,j \in [p], \ell \in [n_{i,j}]} \mathbb{E}(|\zeta_{i,j,\ell}|^4) \leq C$. It follows from Davydov's inequality that $\sum_{\ell_1, \ell_2=1}^{n_{i,j}} |\text{Cov}(\zeta_{i,j,\ell_1}, \zeta_{i,j,\ell_2})| \lesssim \sum_{\ell_2=1}^{n_{i,j}} 1 + \sum_{\ell_2=1}^{n_{i,j}} \sum_{\ell_1=\ell_2+1}^{n_{i,j}} \exp(-C L_{\tilde{n}}^{-\varphi} |\ell_1 - \ell_2|^\varphi) \lesssim n_* L_n$. By Lemma 1 with $m = 0$ and $\tilde{L}_{\tilde{n}} = L_n$, we have

$$\max_{i,j \in [p]} \mathbb{P}(|I_1(i, j)| \geq v) \lesssim (1 + n_* L_n^{-1} \rho^{-1} v^2)^{-\rho/2} + v^{-1} \exp\{-C(n_* L_n^{-1} \rho^{-1} v)^{\varphi/(\varphi+1)}\} \tag{22}$$

for any $v > 0$ and $\rho \geq 1$.

Define $\eta_{i,j,k} = (2N_{i,j,k})^{-1} U_{i,t_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}$. Then $I_2(i, j) = n_{i,j}^{-1} \sum_{k=1}^{n_{i,j}} \eta_{i,j,k}$. Since $\max_{i,j \in [p], k \in [n_{i,j}]} N_{i,j,k} \leq 2\Delta_K + 2$ for sufficiently large n , we know $\max_{i,j \in [p], k \in [n_{i,j}]} \eta_{i,j,k}$ is uniformly bounded away from infinity due to the fact Δ_K is a fixed

constant. It follows from Assumption 3 that

$$\begin{aligned} \max_{i,j \in [p], k \in [n_{ij}]} \mathbb{P}(|\eta_{i,j,k}| \geq x) &\leq \max_{i,j \in [p], k \in [n_{ij}]} \mathbb{P}\left(\left|\frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} U_{j,t_{i,j,\ell}}\right| \geq \sqrt{x}\right) + \max_{i,j \in [p], k \in [n_{ij}]} \mathbb{P}(|U_{i,t_{i,j,k}}| \geq 2\sqrt{x}) \\ &\leq \max_{i,j \in [p], k \in [n_{ij}]} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{P}(|U_{j,t_{i,j,\ell}}| \geq \sqrt{x}) + \max_{i,j \in [p], k \in [n_{ij}]} \mathbb{P}(|U_{i,t_{i,j,k}}| \geq 2\sqrt{x}) \\ &\lesssim \exp(-Cx) \end{aligned}$$

for any $x > 0$. By Davydov's inequality, we have $\sum_{k_1, k_2=1}^{n_{ij}} |\text{Cov}(\eta_{i,j,k_1}, \eta_{i,j,k_2})| \lesssim n_*(K + L_n)$. Write $\hat{\eta}_{i,j,k} = \eta_{i,j,k} - \mu_{i,j,k}$ with $\mu_{i,j,k} = \mathbb{E}(\eta_{i,j,k})$. By Lemma 1 with $m = 2(K + \Delta_K)$ and $\tilde{L}_n = L_n$, $\max_{i,j \in [p]} \mathbb{P}(|n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \hat{\eta}_{i,j,k}| \geq x) \lesssim \{1 + n_*(K + L_n)^{-1} \rho^{-1} x^2\}^{-\rho/2} + x^{-1} \exp\{-C(n_* L_n^{-1} \rho^{-1} x)^{\varphi/(\varphi+1)}\} + x^{-1} \exp(-Cn_* K^{-1} \rho^{-1} x)$ for any $x > 0$ and $\rho \geq 1$. Let $S_{i,j,k,1} = \{t_{i,j,\ell} \in S_{i,j,k} : \ell < k\}$ and $S_{i,j,k,2} = \{t_{i,j,\ell} \in S_{i,j,k} : \ell > k\}$. Then $\mu_{i,j,k} = (2N_{i,j,k})^{-1} \{\mathbb{E}(U_{i,t_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k,1}} U_{j,t_{i,j,\ell}}) + \mathbb{E}(U_{i,t_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k,2}} U_{j,t_{i,j,\ell}})\}$. By Davydov's inequality and Jensen's inequality, it holds that $|\mathbb{E}(U_{i,t_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k,1}} U_{j,t_{i,j,\ell}})| \lesssim \{\mathbb{E}(|U_{i,t_{i,j,k}}|^4)\}^{1/4} \{\mathbb{E}(|\sum_{t_{i,j,\ell} \in S_{i,j,k,1}} U_{j,t_{i,j,\ell}}|^4)\}^{1/4} \alpha_n^{1/2}(K) \lesssim \exp(-CL_n^{-\varphi} K^\varphi)$. Similarly, we have $|\mathbb{E}(U_{i,t_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k,2}} U_{j,t_{i,j,\ell}})| \lesssim \exp(-CL_n^{-\varphi} K^\varphi)$. Thus, it holds that $\max_{i,j \in [p], k \in [n_{ij}]} |\mu_{i,j,k}| \lesssim \exp(-CL_n^{-\varphi} K^\varphi)$, which implies $\max_{i,j \in [p]} |n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \mu_{i,j,k}| \lesssim \exp(-CL_n^{-\varphi} K^\varphi)$. For any $v \gg \exp(-CL_n^{-\varphi} K^\varphi)$ and $\rho \geq 1$, we have

$$\begin{aligned} \max_{i,j \in [p]} \mathbb{P}\{|I_2(i, j)| \geq v\} &\leq \max_{i,j \in [p]} \mathbb{P}\left(\left|\frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \hat{\eta}_{i,j,k}\right| \geq v - \left|\frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \mu_{i,j,k}\right|\right) \leq \max_{i,j \in [p]} \mathbb{P}\left(\left|\frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \hat{\eta}_{i,j,k}\right| \geq \frac{v}{2}\right) \\ &\lesssim \{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp\{-C(n_* L_n^{-1} \rho^{-1} v)^{\varphi/(\varphi+1)}\} + v^{-1} \exp(-Cn_* K^{-1} \rho^{-1} v). \end{aligned}$$

Analogously,

$$\max_{i,j \in [p]} \mathbb{P}\{|I_3(i, j)| \geq v\} \lesssim \{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp\{-C(n_* L_n^{-1} \rho^{-1} v)^{\varphi/(\varphi+1)}\} + v^{-1} \exp(-Cn_* K^{-1} \rho^{-1} v)$$

for any $v \gg \exp(-CL_n^{-\varphi} K^\varphi)$ and $\rho \geq 1$. Note that $I(i, j) = I_1(i, j) + I_2(i, j) + I_3(i, j)$. Together with (22), it holds that

$$\max_{i,j \in [p]} \mathbb{P}\{|I(i, j)| \geq v\} \lesssim \{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp\{-C(n_* L_n^{-1} \rho^{-1} v)^{\varphi/(\varphi+1)}\} + v^{-1} \exp(-Cn_* K^{-1} \rho^{-1} v)$$

for any $v \gg \exp(-CL_n^{-\varphi} K^\varphi)$ and $\rho \geq 1$.

By (20), $|I(i, j)|^m \lesssim |I_1(i, j)|^m + |I_2(i, j)|^m + |I_3(i, j)|^m$ for any fixed positive integer m . By Assumption 3, $\mathbb{E}\{|I_1(i, j)|^m\} \lesssim n_{ij}^{-1} \sum_{\ell=1}^{n_{ij}} \mathbb{E}(|U_{i,t_{i,j,\ell}} U_{j,t_{i,j,\ell}} - \sigma_{u,i,j}|^m) \lesssim 1$ and $\mathbb{E}\{|I_2(i, j)|^m\} \lesssim n_{ij}^{-1} \sum_{k=1}^{n_{ij}} N_{i,j,k}^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}(|U_{i,t_{i,j,k}} U_{j,t_{i,j,\ell}}|^m) \lesssim 1$. Analogously, we also have $\mathbb{E}\{|I_3(i, j)|^m\} \lesssim 1$. Thus, $\max_{i,j \in [p]} \mathbb{E}\{|I(i, j)|^m\} \lesssim 1$. We complete the proof of Proposition 1. \square

8.3. Proof of Proposition 2

Notice that $dX_{i,t} = \mu_{i,t} dt + \sigma_{i,t} dB_{i,t}$. Then

$$\begin{aligned} \mathbb{II}(i, j) &= \underbrace{\frac{1}{2n_{ij}} \sum_{k=1}^{n_{ij}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds\right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s} ds\right)}_{\mathbb{II}_1(i,j)} \\ &+ \underbrace{\frac{1}{2n_{ij}} \sum_{k=1}^{n_{ij}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s}\right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s}\right)}_{\mathbb{II}_2(i,j)} \\ &+ \underbrace{\frac{1}{2n_{ij}} \sum_{k=1}^{n_{ij}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds\right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s}\right)}_{\mathbb{II}_3(i,j)} \\ &+ \underbrace{\frac{1}{2n_{ij}} \sum_{k=1}^{n_{ij}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s}\right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s} ds\right)}_{\mathbb{II}_4(i,j)}. \end{aligned} \tag{23}$$

Recall that $\xi = \max_{i,j \in [p]} \max_{k \in [n_{ij}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} |t_{i,j,\ell} - t_{i,j,k}|$. In the sequel, we will bound the tail probabilities of $\max_{i,j \in [p]} |\mathbb{II}_1(i, j)|$, $\max_{i,j \in [p]} |\mathbb{II}_2(i, j)|$, $\max_{i,j \in [p]} |\mathbb{II}_3(i, j)|$ and $\max_{i,j \in [p]} |\mathbb{II}_4(i, j)|$, respectively.

For any $k = [n_{i,j}]$, define $\zeta_{i,j,k}^* = N_{i,j,k}^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds) (\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s} ds)$. Then we have $\Pi_1(i, j) = (2n_{i,j})^{-1} \sum_{k=1}^{n_{i,j}} \zeta_{i,j,k}^*$. We will first bound $\mathbb{E}\{\exp(\theta \zeta_{i,j,k}^*)\}$ for any $|\theta| \in (0, C_5 \xi^{-2}]$, where C_5 is specified in Assumption 4. By Jensen's inequality and Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}\{\exp(\theta \zeta_{i,j,k}^*)\} &\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E} \left[\exp \left\{ \theta \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s} ds \right) \right\} \right] \\ &\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \frac{1}{|t_{i,j,\ell} - t_{i,j,k}|^2} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \mathbb{E}\{\exp(\theta |t_{i,j,\ell} - t_{i,j,k}|^2 \mu_{i,s_1} \mu_{j,s_2})\} ds_1 ds_2 \\ &\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \frac{1}{|t_{i,j,\ell} - t_{i,j,k}|^2} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} [\mathbb{E}\{\exp(|\theta| |t_{i,j,\ell} - t_{i,j,k}|^2 \mu_{i,s_1}^2)\}]^{1/2} ds_1 \\ &\quad \times \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} [\mathbb{E}\{\exp(|\theta| |t_{i,j,\ell} - t_{i,j,k}|^2 \mu_{j,s_2}^2)\}]^{1/2} ds_2 \\ &\leq \max_{i \in [p]} \sup_{0 \leq s \leq T} \mathbb{E}\{\exp(|\theta| \xi^2 \mu_{i,s}^2)\}. \end{aligned} \tag{24}$$

Recall that $\xi \asymp Kn_*^{-1} = o(1)$. By Assumption 4, $\max_{i \in [p]} \sup_{0 \leq s \leq T} \mathbb{E}\{\exp(|\theta| \xi^2 \mu_{i,s}^2)\} \leq \exp(C_5 C_7) \cdot \exp(C_6 \xi^4 \theta^2) \lesssim \exp(C \xi \theta^2)$ for any $|\theta| \in (0, C_5 \xi^{-2}]$. Therefore, by (24), $\max_{i,j \in [p], k \in [n_{i,j}]} \mathbb{E}\{\exp(\theta \zeta_{i,j,k}^*)\} \lesssim \exp(C \xi \theta^2)$ for any $|\theta| \in (0, C_5 \xi^{-2}]$. By Lemma 2 of Fan et al. (2012), it holds that

$$\max_{i,j \in [p]} \mathbb{P}\{|\Pi_1(i, j)| > v\} \lesssim \exp(-Cv^2 \xi^{-1}) \tag{25}$$

for any $v = o(\xi^{-1})$.

For any $k \in [n_{i,j}]$, define $\zeta_{i,j,k}^{**} = N_{i,j,k}^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s}) (\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s})$. Then we have $\Pi_2(i, j) = (2n_{i,j})^{-1} \sum_{k=1}^{n_{i,j}} \zeta_{i,j,k}^{**}$. For any constant $d \in (0, \xi^{-1/2}]$, define a stopping time $\Gamma_{i,d} = T \wedge \inf\{t > 0 : \sup_{0 \leq s \leq t} \sigma_{i,s} > d\}$. For any $|\theta| \in (0, d^{-2} \xi^{-1}/4]$, by Jensen's inequality and Cauchy-Schwarz inequality, it holds that

$$\begin{aligned} &\mathbb{E}\{\exp(\theta \zeta_{i,j,k}^{**}) I(\Gamma_{i,d} = \Gamma_{j,d} = T)\} \\ &\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E} \left[\exp \left\{ \theta \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s} \right) \right\} I(\Gamma_{i,d} = \Gamma_{j,d} = T) \right] \\ &\leq \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\mathbb{E} \left[\exp \left\{ |\theta| \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right)^2 \right\} I(\Gamma_{i,d} = T) \right] \right)^{1/2} \\ &\quad \times \left(\mathbb{E} \left[\exp \left\{ |\theta| \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s} \right)^2 \right\} I(\Gamma_{j,d} = T) \right] \right)^{1/2}. \end{aligned} \tag{26}$$

Restricted on the event $\{\Gamma_{i,d} = T\}$, we have $\sup_{0 \leq s \leq T} \sigma_{i,s} \leq d$. For any $|\theta| \in (0, d^{-2} \xi^{-1}/4]$, it holds that

$$\begin{aligned} &\exp \left\{ |\theta| \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right)^2 \right\} I(\Gamma_{i,d} = T) \\ &\leq \exp \left[|\theta| \left\{ \left(\int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right)^2 - \int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{i,s}^2 ds \right\} \right] I(\Gamma_{i,d} = T) \\ &\quad \times \exp \left(|\theta| \int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{i,s}^2 ds \right) I(\Gamma_{i,d} = T) \\ &\leq C \exp \left[|\theta| \left\{ \left(\int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right)^2 - \int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{i,s}^2 ds \right\} \right] I(\Gamma_{i,d} = T). \end{aligned} \tag{27}$$

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

$$\begin{aligned} &\mathbb{E} \left\{ \exp \left[|\theta| \left\{ \left(\int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right)^2 - \int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{i,s}^2 ds \right\} \right] I(\Gamma_{i,d} = T) \middle| \mathcal{F}_{i,t_{i,j,\ell} \wedge t_{i,j,k}}^* \right\} \\ &\leq \mathbb{E}[\exp\{|\theta| (B_{i,d^2|t_{i,j,\ell} - t_{i,j,k}|}^2 - d^2 |t_{i,j,\ell} - t_{i,j,k}|)\}] = \mathbb{E}[\exp\{|\theta| d^2 |t_{i,j,\ell} - t_{i,j,k}| (Z^2 - 1)\}] \\ &\leq \exp(Cd^4 \xi^2 \theta^2) \leq \exp(C(d \xi^{1/2})^4 \theta^2) \end{aligned} \tag{28}$$

for any $\tau \in [0, 4)$, where $Z \sim N(0, 1)$ and $\mathcal{F}_{i,t}^*$ is the σ -field generated by $(\sigma_{i,s}, B_{i,s})_{0 \leq s \leq t}$. Thus, by (26) and (27), we have $\max_{i,j \in [p], k \in [n_{i,j}]} \mathbb{E}\{\exp(\theta \zeta_{i,j,k}^{**}) I(\Gamma_{i,d} = \Gamma_{j,d} = T)\} \lesssim \exp\{C(d\xi^{1/2})^{4-\tau} \theta^2\}$ for any $|\theta| \in (0, d^{-2} \xi^{-1}/4]$. By Lemma 2 of Fan et al. (2012), it holds that $\max_{i,j \in [p]} \mathbb{P}\{|\Pi_2(i, j)| > v, \Gamma_{i,d} = \Gamma_{j,d} = T\} \lesssim \exp\{-Cv^2(d\xi^{1/2})^{\tau-4}\}$ for any $v = o\{(d\xi^{1/2})^{2-\tau}\}$. Note that

$$\mathbb{P}\{|\Pi_2(i, j)| > v\} \leq \mathbb{P}\{|\Pi_2(i, j)| > v, \Gamma_{i,d} = \Gamma_{j,d} = T\} + \mathbb{P}(\Gamma_{i,d} \neq T) + \mathbb{P}(\Gamma_{j,d} \neq T).$$

Since $\Gamma_{i,d} = T \wedge \inf\{t > 0 : \sup_{0 \leq s \leq t} \sigma_{i,s} > d\}$, by Assumption 5, we have $\max_{i \in [p]} \mathbb{P}(\Gamma_{i,d} \neq T) \leq \max_{i \in [p]} \mathbb{P}(\sup_{0 \leq s \leq T} \sigma_{i,s} > d) \lesssim \exp(-Cd^\gamma)$. Then

$$\max_{i,j \in [p]} \mathbb{P}\{|\Pi_2(i, j)| > v\} \lesssim \exp\{-Cv^2(d\xi^{1/2})^{\tau-4}\} + \exp(-Cd^\gamma) \tag{29}$$

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

Due to $xy \leq 2^{-1}(x^2 + y^2)$ for any $x, y > 0$, we have

$$2|\Pi_3(i, j)| \leq \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds \right)^2 + \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s} \right)^2.$$

Identical to deriving (25) and (29), it holds that

$$\max_{i,j \in [p]} \mathbb{P}\{|\Pi_3(i, j)| > v\} \lesssim \exp(-Cv^2\xi^{-1}) + \exp\{-Cv^2(d\xi^{1/2})^{\tau-4}\} + \exp(-Cd^\gamma)$$

for any $v = o\{(d\xi^{1/2})^{2-\tau}\}$ with $d \leq \xi^{-1/2}$. Such upper bound also holds for $\max_{i,j \in [p]} \mathbb{P}\{|\Pi_4(i, j)| > v\}$. Together with (25) and (29), we have

$$\max_{i,j \in [p]} \mathbb{P}\{|\Pi(i, j)| > v\} \lesssim \exp(-Cv^2\xi^{-1}) + \exp\{-Cv^2(d\xi^{1/2})^{\tau-4}\} + \exp(-Cd^\gamma) \tag{30}$$

for any $v = o\{(d\xi^{1/2})^{2-\tau} \wedge \xi^{-1}\}$ with $d \leq \xi^{-1/2}$. To make $\max_{i,j \in [p]} |\Pi(i, j)| = O_p(n_*^{-1/2} K^{1/2} \log^{1/2} p)$, it suffices to require $d^{2\tau-4} \xi^{\tau-1} \log p = o(1)$, $\xi^3 \log p = o(1)$, $d^{8-2\tau} \xi^{2-\tau} = O(1)$, $d^2 \xi \leq 1$ and $\log p = o(d^\gamma)$. Due to $\tau \in [0, 4)$, $\xi = o(1)$ and $d \rightarrow \infty$, $d^{8-2\tau} \xi^{2-\tau} \rightarrow \infty$ when $\tau \in [2, 4)$. Thus, we need to restrict $\tau \in [0, 2)$, which leads to $\xi^{(\tau-1)/(2-\tau)} (\log p)^{1/(2-\tau)} \ll d^2 \ll \xi^{(\tau-2)/(4-\tau)}$ and $\log p = o(d^\gamma \wedge \xi^{-3})$. It follows from $\xi^{(\tau-1)/(2-\tau)} (\log p)^{1/(2-\tau)} \ll \xi^{(\tau-2)/(4-\tau)}$ that $\xi^{\tau/(4-\tau)} \log p = o(1)$. Selecting d sufficiently close to $\xi^{(\tau-2)/(8-2\tau)}$, we have $\log p = o[\min\{\xi^{-\tau/(4-\tau)}, \xi^{\gamma(\tau-2)/(8-2\tau)}\}]$. To make p diverge as fast as possible, we can choose $\tau = 2\gamma/(\gamma + 2)$ and $d = \xi^{-1/(4+\gamma)}$. Note that $\xi^{-1} \asymp n_* K^{-1}$. It follows from (30) that

$$\max_{i,j \in [p]} \mathbb{P}\{|\Pi(i, j)| > v\} \lesssim \exp(-Cn_* K^{-1} v^2) + \exp\{-C(n_* K^{-1})^{\gamma/(\gamma+4)} v\}$$

for any $v = o\{(n_*^{-1} K)^{2/(\gamma+4)}\}$.

By (23), $\Pi(i, j) = \Pi_1(i, j) + \Pi_2(i, j) + \Pi_3(i, j) + \Pi_4(i, j)$. Notice that for sufficiently large n , $\min_{k \in [n_{i,j}]} N_{i,j,k} = \Delta_K + 1$ and $\max_{k \in [n_{i,j}]} N_{i,j,k} = 2\Delta_K + 2$. Since Δ_K is a fixed constant, by Assumption 4, for any fixed positive integer m , Jensen's inequality implies that

$$\begin{aligned} \max_{i,j \in [p]} \mathbb{E}\{|\Pi_1(i, j)|^m\} &\lesssim \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\left\{ \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds \right)^{2m} + \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s} ds \right)^{2m} \right\} \\ &\lesssim \xi^{2m-1} \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sup_{0 \leq s \leq T} \mathbb{E}\{|\mu_{i,s}|^{2m}\} ds \lesssim \xi^{2m}. \end{aligned} \tag{31}$$

Meanwhile, by Assumption 4 and Burkholder–Davis–Gundy inequality, it holds that

$$\max_{i,j \in [p]} \mathbb{E}\{|\Pi_2(i, j)|^m\} \lesssim \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\left\{ \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right)^{2m} \right\} \lesssim \xi^m. \tag{32}$$

Due to $\max_{i,j \in [p]} \mathbb{E}\{|\Pi_3(i, j)|^m + |\Pi_4(i, j)|^m\} \lesssim \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\{(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds)^{2m} + (\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s})^{2m}\}$, together with (31) and (32), $\max_{i,j \in [p]} \mathbb{E}\{|\Pi_3(i, j)|^m + |\Pi_4(i, j)|^m\} \lesssim \xi^m$. Hence, $\max_{i,j \in [p]} \mathbb{E}\{|\Pi(i, j)|^m\} \lesssim \xi^m$. We complete the proof of Proposition 2. \square

8.4. Proof of Proposition 3

To prove Proposition 3, we need the following lemma whose proof is given in Section 8.17.

Lemma 2. Let $\{z_t\}_{t=1}^{\tilde{n}}$ be an α -mixing sequence of real-valued and centered random variables with α -mixing coefficients $\{\alpha(k)\}_{k \geq 1}$. Assume there exist some universal constants $a_1 > 1$, $a_2 > 0$, $\tilde{\tau} \geq 0$ and $\varphi > 0$ such that (i) $\max_{t \in [\tilde{n}]} \mathbb{E}\{|z_t|^k\} \leq$

$(k!)^{1+\tilde{r}} H_{\tilde{n}}^k$ for any integer $k \geq 2$, where $H_{\tilde{n}} > 0$ may diverge with \tilde{n} , (ii) $\alpha(k) \leq a_1 \exp\{-a_2(\tilde{L}_{\tilde{n}}^{-1}k)^\varphi\}$ for any integer $k \geq 1$, where $\tilde{L}_{\tilde{n}} > 0$ may diverge with \tilde{n} . Let $S_{\tilde{n}} = \sum_{t=1}^{\tilde{n}} z_t$. It holds that

$$\mathbb{P}(|S_{\tilde{n}}| \geq \tilde{n}x) \lesssim \exp(-C\tilde{n}\tilde{L}_{\tilde{n}}^{-1}H_{\tilde{n}}^{-2}x^2) + \exp\{-C(\tilde{n}\tilde{L}_{\tilde{n}}^{-1}H_{\tilde{n}}^{-1}x)^{1/(1+\tilde{r})}\}$$

for any $x > 0$, where $\check{r} = 1 + \tilde{r} + \varphi^{-1}$.

Remark 8. If $\varphi = \infty$, the upper bound in Lemma 2 holds with $\check{r} = 1 + \tilde{r}$.

By the definition of III(i, j), we can reformulate it as

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

For each $i, j \in [p]$ and $k \in [n_{i,j}]$, define

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

and $D_{i,j} = \max_{k \in [n_{i,j}]} |G_{i,j,k}|$. Recall $\xi = \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} |t_{i,j,\ell} - t_{i,j,k}|$.

We will first consider the tail probability $\mathbb{P}(D_{i,j} > v)$. By Bonferroni inequality, we have

$$\begin{aligned} \mathbb{P}(D_{i,j} > v) &\leq \sum_{k=1}^{n_{i,j}} \mathbb{P} \left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) (X_{i,t_{i,j,k}} - X_{i,t_{i,j,\ell}}) \right| > v \right\} \\ &\leq \sum_{k=1}^{n_{i,j}} \mathbb{P} \left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s} ds \right| > \frac{v}{2} \right\} \\ &\quad + \sum_{k=1}^{n_{i,j}} \mathbb{P} \left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right| > \frac{v}{2} \right\}. \end{aligned} \tag{33}$$

For any $\theta > 0$, by Triangle inequality and Jensen's inequality, similar to (24), it holds that

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

It follows from Assumption 4 that $\sup_{0 \leq t \leq T} \mathbb{E} \{ \exp(C\theta \xi |\mu_{i,t}|) \} \leq C \exp(C\xi^2\theta^2)$. Selecting $\theta \asymp \xi^{-1}$ and applying Markov's inequality, we have

$$\begin{aligned} \max_{k \in [n_{i,j}]} \mathbb{P} \left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s} ds \right| > \frac{v}{2} \right\} \\ \leq \exp(-\theta v) \max_{k \in [n_{i,j}]} \mathbb{E} \left[\exp \left\{ 2\theta \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s} ds \right| \right\} \right] \\ \lesssim \exp(-C\xi^{-1}v) \end{aligned} \tag{34}$$

for any $v > 0$. For any constant $d \in (0, \xi^{-1/2}]$, define a stopping time $\Gamma_{i,d} = T \wedge \inf\{t > 0 : \sup_{0 \leq s \leq t} \sigma_{i,s} > d\}$. By Cauchy-Schwarz inequality,

$$\left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right|^2 \lesssim \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right)^2,$$

which implies that for any $\theta > 0$,

$$\begin{aligned} \mathbb{P} \left\{ \left| \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}} \right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right| > \frac{v}{2}, \Gamma_{i,d} = T \right\} \\ \lesssim \exp(-C\theta v^2) \mathbb{E} \left[\exp \left\{ \frac{\theta}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s} \right)^2 \right\} I(\Gamma_{i,d} = T) \right]. \end{aligned} \tag{35}$$

By Jensen's inequality, $\mathbb{E}[\exp\{\theta N_{i,j,k}^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (\int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s})^2\} | (\Gamma_{i,d} = T)] \leq N_{i,j,k}^{-1} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}[\exp\{\theta (\int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s})^2\} | (\Gamma_{i,d} = T)]$. Same as (27) and (28), for any $\theta \in (0, d^{-2}\xi^{-1}/4]$, $\mathbb{E}[\exp\{\theta (\int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s})^2\} | (\Gamma_{i,d} = T)] \leq \exp(Cd^4\xi^2\theta^2) \lesssim 1$. Selecting $\theta = d^{-2}\xi^{-1}/4$, together with (35), we have $\max_{k \in [n_{i,j}]} \mathbb{P}\{|\sum_{t_{i,j,\ell} \in S_{i,j,k}} (N_{i,j,\ell}^{-1} + N_{i,j,k}^{-1}) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s}| > v/2, \Gamma_{i,d} = T\} \lesssim \exp(-Cd^{-2}\xi^{-1}v^2)$ for any $v > 0$. It follows from Assumption 5 that

$$\begin{aligned} & \max_{k \in [n_{i,j}]} \mathbb{P}\left\{\left|\sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}}\right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s}\right| > \frac{v}{2}\right\} \\ & \leq \max_{k \in [n_{i,j}]} \mathbb{P}\left\{\left|\sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\frac{1}{N_{i,j,\ell}} + \frac{1}{N_{i,j,k}}\right) \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s}\right| > \frac{v}{2}, \Gamma_{i,d} = T\right\} + \mathbb{P}(\Gamma_{i,d} \neq T) \\ & \lesssim \exp(-Cd^\gamma) + \exp(-Cd^{-2}\xi^{-1}v^2) \end{aligned}$$

for any $v > 0$. Letting $d \rightarrow \infty$, together with (34), (33) implies that

$$\max_{i,j \in [p]} \mathbb{P}(D_{i,j} > v) \lesssim n_* \exp(-Cd^\gamma) + n_* \exp(-Cd^{-2}\xi^{-1}v^2) \tag{36}$$

for any $0 < v \leq C$. Write $\bar{\mathbb{E}}(\cdot) = \mathbb{E}\{\cdot \mid (\mathbf{X}_t)_{t \in [0,T]}\}$. For any integer $s \geq 2$, we have $\bar{\mathbb{E}}(|G_{i,j,k} U_{j,t_{i,j,k}}|^s) \leq D_{i,j}^s \cdot \mathbb{E}(|U_{j,t_{i,j,k}}|^s) \leq D_{i,j}^s \cdot C^s s^{(s+1)/2} \leq s! \cdot (CD_{i,j})^s$. Applying Lemma 2 with $\tilde{L}_{\bar{n}} = L_n$, $\mathbb{P}\{|\text{III}(i,j)| \geq x \mid (\mathbf{X}_t)_{t \in [0,T]}\} \lesssim \exp(-Cn_*L_n^{-1}D_{i,j}^{-2}x^2) + \exp\{-C(n_*L_n^{-1}D_{i,j}^{-1}x)^{\varphi/(2\varphi+1)}\}$ for any $x > 0$, which implies that

$$\mathbb{P}\{|\text{III}(i,j)| \geq x\} \lesssim \mathbb{E}\left\{\exp\left(-\frac{Cn_*L_n^{-1}x^2}{D_{i,j}^2}\right)\right\} + \mathbb{E}\left[\exp\left\{-C\left(\frac{n_*L_n^{-1}x}{D_{i,j}}\right)^{\varphi/(2\varphi+1)}\right\}\right]$$

for any $x > 0$. Therefore, from (36) with $v \asymp 1$, it holds that

$$\begin{aligned} \mathbb{E}\left\{\exp\left(-\frac{Cn_*L_n^{-1}x^2}{D_{i,j}^2}\right)\right\} & \lesssim \exp(-Cv^{-2}n_*L_n^{-1}x^2) + \mathbb{P}(D_{i,j} > v) \\ & \lesssim \exp(-Cn_*L_n^{-1}x^2) + n_* \exp(-Cd^\gamma) + n_* \exp(-Cd^{-2}\xi^{-1}). \end{aligned}$$

Analogously, we also have

$$\mathbb{E}\left[\exp\left\{-C\left(\frac{n_*L_n^{-1}x}{D_{i,j}}\right)^{\varphi/(2\varphi+1)}\right\}\right] \lesssim \exp\{-C(n_*L_n^{-1}x)^{\varphi/(2\varphi+1)}\} + n_* \exp(-Cd^\gamma) + n_* \exp(-Cd^{-2}\xi^{-1}).$$

Thus, $\max_{i,j \in [p]} \mathbb{P}\{|\text{III}(i,j)| \geq x\} \lesssim \exp(-Cn_*L_n^{-1}x^2) + \exp\{-C(n_*L_n^{-1}x)^{\varphi/(2\varphi+1)}\} + n_* \exp(-Cd^\gamma) + n_* \exp(-Cd^{-2}\xi^{-1})$ for any $x > 0$. Recall that $\xi \asymp Kn_*^{-1}$. To make $\max_{i,j \in [p]} |\text{III}(i,j)| = O_p(n_*^{-1/2}K^{1/2} \log^{1/2} p)$, it suffices to require $K \gtrsim L_n$, $\log p = o[\min\{d^\gamma, d^{-2}n_*K^{-1}, (n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}\}]$ and $\log n_* = o[\min\{d^\gamma, d^{-2}n_*K^{-1}\}]$. In order to make p diverge as fast as possible, we can select $d = (n_*K^{-1})^{1/(2+\gamma)}$. If $K(\log n_*)^{1+2/\gamma} = o(n_*)$, then $\max_{i,j \in [p]} \mathbb{P}\{|\text{III}(i,j)| \geq v\} \lesssim \exp(-Cn_*L_n^{-1}v^2) + \exp\{-C(n_*K^{-1})^{\gamma/(2+\gamma)}\}$ for any $v > 0$.

By Jensen's inequality and Cauchy-Schwarz inequality, it holds that

$$\begin{aligned} \mathbb{E}\{|\text{III}(i,j)|^m\} & \lesssim \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}(|X_{i,t_{i,j,k}} - X_{i,t_{i,j,\ell}}|^m |U_{j,t_{i,j,k}}|^m) \\ & = \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\left(\left|\int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s} ds + \int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s}\right|^m |U_{j,t_{i,j,k}}|^m\right) \\ & \lesssim \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \left\{\mathbb{E}\left(\left|\int_{t_{i,j,\ell}}^{t_{i,j,k}} \mu_{i,s} ds\right|^{2m} + \left|\int_{t_{i,j,\ell}}^{t_{i,j,k}} \sigma_{i,s} dB_{i,s}\right|^{2m}\right)\right\}^{1/2} \{\mathbb{E}(U_{j,t_{i,j,k}}^{2m})\}^{1/2}. \end{aligned}$$

By Assumption 3, $\max_{k \in [n_{i,j}]} \mathbb{E}(U_{j,t_{i,j,k}}^{2m}) \lesssim 1$. Together with (31) and (32), $\max_{i,j \in [p]} \mathbb{E}\{|\text{III}(i,j)|^m\} \lesssim \xi^{m/2}$. Analogously, we can show the same results hold for $\text{IV}(i,j)$. We complete the proof of Proposition 3. \square

8.5. Proof of Theorem 2

To prove Theorem 2, we need the Le Cam's lemma as stated in Lemma 3. 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}_θ 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 $\Theta = \{\theta_0, \theta_1, \dots, \theta_D\}$ and denote by L the loss function. Define $l_{\min} = \min_{d \in [D]} \inf_t \{L(t, \theta_0) + L(t, \theta_d)\}$ and denote $\bar{\mathbb{P}} = D^{-1} \sum_{d=1}^D \mathbb{P}_{\theta_d}$.

Lemma 3 (Le Cam's lemma). Let T be any estimator of θ based on an observation \mathcal{Z} from a distribution \mathbb{P}_θ with $\theta \in \Theta = \{\theta_0, \theta_1, \dots, \theta_D\}$, then $\sup_{\theta \in \Theta} \mathbb{E}_{\mathcal{Z}|\theta} \{L(T, \theta)\} \geq 2^{-1} l_{\min} \|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\|$.

For each $k \in [n]$, define $C_k = \{i \in [p] : t_k \in \mathcal{G}_i\}$ where \mathcal{G}_i is the grid of time points where we observe the noisy data of the i th component process. For any s -dimensional vector \mathbf{a} and an index set $\mathcal{C} \subset [s]$, denote by $\mathbf{a}_{\mathcal{C}}$ the subvector of \mathbf{a} with components indexed by \mathcal{C} . The data we have is $\mathcal{Z} = \{\mathbf{Y}_{t_1, C_1}, \dots, \mathbf{Y}_{t_n, C_n}\}$. Select the loss function $L(T, \theta) = \max_{i,j \in [p]} |\omega_{i,j} - \theta_{i,j}|$ for any $T = (\omega_{i,j})_{p \times p}$ and $\theta = (\theta_{i,j})_{p \times p} \in \Theta$. Select $D = p$, $\theta_0 = \Sigma_{u,0} = \mathbf{I}_p$ and

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

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

$$\begin{aligned} l_{\min} &= \min_{d \in [D]} \inf_t \{L(t, \theta_0) + L(t, \theta_d)\} \\ &\geq \min_{d \in [D]} \max_{i,j \in [p]} |\sigma_{u,i,j,0} - \sigma_{u,i,j,d}| \geq \sqrt{\frac{v \log p}{n_*}} \asymp \sqrt{\frac{\log p}{n_*}}. \end{aligned} \tag{37}$$

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 $n \geq n_*$, and $0 \leq t_1 < \dots < t_n = T$, we define $\mathcal{G}_* = \{\tilde{t}_1, \dots, \tilde{t}_{n_*}\}$ with each $\tilde{t}_j \in \{t_1, \dots, t_n\}$ and $\tilde{t}_j < \tilde{t}_{j+1}$. For each $\tilde{t}_j \in \mathcal{G}_*$, we assume all p component processes are observed. For any $t_j \notin \mathcal{G}_*$, we assume only one component process are observed. Without loss of generality, we assume $\mathcal{G}_* = \{t_1, \dots, t_{n_*}\}$. Let $n - n_* = ap + q_*$ where $a \geq 0$ and $0 \leq q_* < p$ are two integers. We assume the i th component process is observed at t_{n_*+jp+i} 's with $j = 0, \dots, a$ and $i \in [p]$. Then $\mathcal{G}_i = \mathcal{G}_* \cup \{t_{n_*+i}, \dots, t_{n_*+ap+i}\}$.

Let $\mathbf{U}_{t_k} \stackrel{i.i.d.}{\sim} N(\mathbf{0}, \Sigma_{u,d})$, and denote the joint density of $\mathbf{U}_{t_1, C_1}, \dots, \mathbf{U}_{t_n, C_n}$ 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 $f_0 = \prod_{k=1}^n \prod_{j \in C_k} \phi_1(u_{k,j})$ and $f_d = \prod_{k=1}^n \prod_{j \in C_k \setminus \{d\}} \phi_1(u_{k,j}) \cdot \prod_{k=1}^n \prod_{d \in C_k} \phi_{\sigma_*}(u_{k,d})$ for each $d \in [D]$. Here we adopt the convention $\prod_{d \in C_k} \phi_{\sigma_*}(u_{k,d}) \equiv 1$ if $d \notin C_k$. We will show $\|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\| \geq c$ for some universal constant $c > 0$.

For any two densities q_0 and q_1 , by Cauchy–Schwarz inequality, we have $(\int |q_0 - q_1| d\mu)^2 \leq \int (q_0 - q_1)^2 / q_1 d\mu = \int q_0^2 / q_1 d\mu - 1$, which implies that $\int q_0 \wedge q_1 d\mu = 1 - 2^{-1} \int |q_0 - q_1| d\mu \geq 1 - 2^{-1} (\int q_0^2 / q_1 d\mu - 1)^{1/2}$. In order to show $\|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\| \geq c$ for some universal constant $c > 0$, it suffices to show that $\int (D^{-1} \sum_{d=1}^D f_d)^2 f_0^{-1} d\mu - 1 \rightarrow 0$, that is,

$$\frac{1}{D^2} \sum_{d=1}^D \left(\int \frac{f_d^2}{f_0} d\mu - 1 \right) + \frac{1}{D^2} \sum_{d_1 \neq d_2} \left(\int \frac{f_{d_1} f_{d_2}}{f_0} d\mu - 1 \right) \rightarrow 0. \tag{38}$$

Since $f_{d_1} f_{d_2} / f_0 = \prod_{k=1}^n [\prod_{d_1 \in C_k} \phi_{\sigma_*}(u_{k,d_1}) \cdot \prod_{d_2 \in C_k} \phi_{\sigma_*}(u_{k,d_2}) \cdot \prod_{j \in C_k \setminus \{d_1, d_2\}} \phi_1(u_{k,j})]$ for any $d_1 \neq d_2$, then $\int f_{d_1} f_{d_2} / f_0 d\mu = 1$, which implies $D^{-2} \sum_{d_1 \neq d_2} (\int f_{d_1} f_{d_2} / f_0 d\mu - 1) = 0$. For any $d \in [D]$, we have

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

which implies

$$\begin{aligned} \int \frac{f_d^2}{f_0} d\mu &= \left(\frac{1}{\sigma_* \sqrt{2 - \sigma_*^2}} \right)^{\sum_{k=1}^n I(d \in C_k)} \prod_{k=1}^n \prod_{j \in C_k \setminus \{d\}} \left\{ \int \phi_1(u_{k,j}) du_{k,j} \right\} \\ &\quad \times \prod_{k=1}^n \left[\int \frac{\sqrt{2 - \sigma_*^2}}{\sqrt{2\pi\sigma_*^2}} \exp\left\{-\frac{(2 - \sigma_*^2)u_{k,d}^2}{2\sigma_*^2}\right\} du_{k,d} \right]^{I(d \in C_k)} \\ &= \left(\frac{1}{\sigma_* \sqrt{2 - \sigma_*^2}} \right)^{\sum_{k=1}^n I(d \in C_k)} = \left(1 - \frac{v \log D}{n_*} \right)^{-\sum_{k=1}^n I(d \in C_k)/2}. \end{aligned}$$

Notice that $\sum_{k=1}^n I(d \in C_k) \leq n_* + a + 1$ for each $d \in [D]$. Therefore, $\int f_d^2 / f_0 d\mu \leq (1 - v n_*^{-1} \log D)^{-(n_*+a+1)/2}$ for each $d \in [D]$. Due to $n/n_* \lesssim p$, we know $a \lesssim n_*$. Applying the inequality $\log(1 - x) \geq -2x$ for any $0 < x < 1/2$, we have

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

for sufficiently small $v > 0$. Then (38) holds. Hence $\|\mathbb{P}_{\theta_0} \wedge \bar{\mathbb{P}}\| \geq c$ for some universal constant $c > 0$. Together with (37), we can obtain Theorem 2 by Lemma 3. \square

8.6. Proof of Theorem 3

We first consider the case with $\varphi < \infty$. Write $\aleph = (n_*^{-1}K \log p)^{1/2}$. For each $i, j \in [p]$, we define the event $A_{i,j} = \{|\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}| \leq 4 \min(|\sigma_{u,i,j}|, \alpha \aleph)\}$ with some $\alpha > 0$, and $d_{i,j} = (\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j})I(A_{i,j}^c)$. Write $\mathbf{D} = (d_{i,j})_{p \times p}$. Due to $\|\mathbf{W}\|_2 \leq \|\mathbf{W}\|_\infty$ for any $p \times p$ symmetric matrix \mathbf{W} , it holds that

$$\begin{aligned} \|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2 &\leq \left\{ \max_{i \in [p]} \sum_{j=1}^p |\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}| \right\}^2 \\ &\leq 2 \left(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}| \right)^2 + 2 \left\{ \max_{i \in [p]} \sum_{j=1}^p |\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}| I(A_{i,j}) \right\}^2. \end{aligned} \tag{39}$$

For the second term on the right-hand side of (39), we have that $\sum_{j=1}^p |\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}| I(A_{i,j}) \leq 4 \sum_{j=1}^p \alpha \aleph I(|\sigma_{u,i,j}| > \alpha \aleph) + 4 \sum_{j=1}^p |\sigma_{u,i,j}| I(|\sigma_{u,i,j}| \leq \alpha \aleph)$. Due to $\sum_{j=1}^p |\sigma_{u,i,j}|^q \leq c_p$, we then have $\sum_{j=1}^p |\sigma_{u,i,j}| I(|\sigma_{u,i,j}| \leq \alpha \aleph) \leq \sum_{j=1}^p |\sigma_{u,i,j}|^q (\alpha \aleph)^{1-q} \leq \alpha^{1-q} c_p \aleph^{1-q}$ and $\sum_{j=1}^p \alpha \aleph I(|\sigma_{u,i,j}| > \alpha \aleph) \leq \sum_{j=1}^p |\sigma_{u,i,j}|^q (\alpha \aleph)^{1-q} \leq \alpha^{1-q} c_p \aleph^{1-q}$. Therefore, we have that $\sum_{j=1}^p |\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}| I(A_{i,j}) \leq 8 \alpha^{1-q} c_p \aleph^{1-q}$ holds uniformly over $i \in [p]$. It follows from (39) that

$$\mathbb{E}(\|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim \mathbb{E} \left\{ \left(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}| \right)^2 \right\} + c_p^2 \aleph^{2(1-q)}. \tag{40}$$

Recall $\hat{\sigma}_{u,i,j}$ is defined as (5). It holds that

$$\begin{aligned} \mathbb{E} \left\{ \left(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}| \right)^2 \right\} &\leq p \sum_{i,j=1}^p \mathbb{E} \{ |\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}|^2 I(A_{i,j}^c) \} \\ &= p \underbrace{\sum_{i,j=1}^p \mathbb{E} (|\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}|^2 I[A_{i,j}^c \cap \{\hat{\sigma}_{u,i,j}^{\text{thre}} = 0\}])}_I \\ &\quad + p \underbrace{\sum_{i,j=1}^p \mathbb{E} (|\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}|^2 I[A_{i,j}^c \cap \{\hat{\sigma}_{u,i,j}^{\text{thre}} = \hat{\sigma}_{u,i,j}\}])}_II. \end{aligned} \tag{41}$$

Recall $\hat{\sigma}_{u,i,j}^{\text{thre}} = \hat{\sigma}_{u,i,j} I(|\hat{\sigma}_{u,i,j}| \geq \beta \aleph)$ for any $i, j \in [p]$. Then

$$\begin{aligned} I &= p \sum_{i,j=1}^p \sigma_{u,i,j}^2 \mathbb{P}[\{|\sigma_{u,i,j}| \geq 4\alpha \aleph\} \cap \{|\hat{\sigma}_{u,i,j}| < \beta \aleph\}] \\ &\leq p \sum_{i,j=1}^p \sigma_{u,i,j}^2 \mathbb{P}[\{|\sigma_{u,i,j}| \geq 4\alpha \aleph\} \cap \{|\sigma_{u,i,j}| - |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| < \beta \aleph\}] \\ &\leq p \sum_{i,j=1}^p \sigma_{u,i,j}^2 \mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| \geq (4\alpha - \beta) \aleph\}. \end{aligned} \tag{42}$$

Selecting $\alpha = \beta/2$ and β being sufficiently large, identical to the arguments used in Section 8.1 for bounding the convergence rate of A_2 , we have $I \lesssim o(\aleph^4)$ provided that $K(\log n_*)^{1+2/\gamma} = o(n_*)$, $K \gtrsim L_n$, $K^{-\varphi} L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, $\log p = o[\min\{(n_* L_n^{-2} K)^{\varphi/(3\varphi+2)}, (n_* K^{-1})^\chi\}]$ with $\chi = \min\{\gamma/(\gamma+4), 1/3\}$. Also, by Cauchy-Schwarz inequality, it holds that

$$\begin{aligned} II &= p \sum_{i,j=1}^p \mathbb{E} \{ |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}|^2 I(|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > 4\alpha \aleph) I(|\sigma_{u,i,j}| > \alpha \aleph) I(|\hat{\sigma}_{u,i,j}| \geq \beta \aleph) \} \\ &\quad + p \sum_{i,j=1}^p \mathbb{E} \{ |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}|^2 I(|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > 4|\sigma_{u,i,j}|) I(|\sigma_{u,i,j}| \leq \alpha \aleph) I(|\hat{\sigma}_{u,i,j}| \geq \beta \aleph) \} \\ &\leq p \sum_{i,j=1}^p \left\{ \mathbb{E}(|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}|^4) \right\}^{1/2} \left\{ \mathbb{P}(|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > 4\alpha \aleph) \right\}^{1/2} \end{aligned} \tag{43}$$

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

Notice that $\max_{i,j \in [p]} \mathbb{E}(|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}|^4) \lesssim 1$. Since $\alpha = \beta/2$, repeating the arguments used in Section 8.1 for bounding the convergence rate of A_2 again, we have

$$p^3 \left\{ \max_{i,j \in [p]} \mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > 4\alpha\mathfrak{N}\} \right\}^{1/2} \leq p^3 \left[\max_{i,j \in [p]} \mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > (\beta - \alpha)\mathfrak{N}\} \right]^{1/2} \lesssim o(\mathfrak{N}^2)$$

provided that $K(\log n_*)^{1+2/\gamma} = o(n_*)$, $K \gtrsim L_n$, $K^{-\varphi} L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$ and $\log p = o[\min\{(n_* L^{-2} K)^\varphi / (3\varphi + 2), (n_* K^{-1})^\chi\}]$. With sufficiently large β , we have $\Pi \lesssim o(\mathfrak{N}^2)$. Together with $I \lesssim o(\mathfrak{N}^4)$, we have $\mathbb{E}\{(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|)^2\} \leq I + \Pi \lesssim o(\mathfrak{N}^2)$. It follows from (40) that $\sup_{\mathcal{P}_2} \mathbb{E}(\|\widehat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim c_p^2 (n_*^{-1} K \log p)^{1-q}$. Analogously, in the case with $\varphi = \infty$, we have $\sup_{\mathcal{P}_2} \mathbb{E}(\|\widehat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim c_p^2 (n_*^{-1} K \log p)^{1-q}$ provided that $K(\log n_*)^{1+2/\gamma} = o(n_*)$, $K > L_n$ and $\log p = o\{(n_* K^{-1})^\chi\}$. We complete the proof of Theorem 3. \square

8.7. 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 $n \geq n_*$, and $0 \leq t_1 < \dots < t_n = T$, we define $\mathcal{G}_* = \{\tilde{t}_1, \dots, \tilde{t}_{n_*}\}$ with each $\tilde{t}_j \in \{t_1, \dots, t_n\}$ 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 loss of generality, we assume $\mathcal{G}_* = \{t_1, \dots, t_{n_*}\}$. Let $n - n_* = ap + q_*$ where $a \geq 0$ and $0 \leq q_* < p$ are two integers. We assume the i th component process is observed at t_{n_*+jp+i} 's with $j = 0, \dots, a$ and $i \in [p]$. Then $\mathcal{G}_i = \mathcal{G}_* \cup \{t_{n_*+i}, \dots, t_{n_*+ap+i}\}$. The data we have is $\mathcal{Z} = \{\mathbf{Y}_{t_1, C_1}, \dots, \mathbf{Y}_{t_n, C_n}\}$ where $C_k = \{i \in [p] : t_k \in \mathcal{G}_i\}$.

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 $v = (v_1, \dots, v_p)$ such that $v_j = 0$ for $1 \leq j \leq p-r$ and $v_j = 0$ or 1 for $p-r+1 \leq j \leq p$ under the restriction $\sum_{j=1}^p |v_j| = K_*$. We will specify K_* later. If each $\lambda_j \in \mathcal{B}$, we say $\lambda = (\lambda_1, \dots, \lambda_r) \in \mathcal{B}^r$. Set $\Gamma = \{0, 1\}^r$ and $\Lambda \subset \mathcal{B}^r$. For each $\lambda = (\lambda_1, \dots, \lambda_r) \in \Lambda$, we define $p \times p$ symmetric matrices $\mathbf{A}_1(\lambda_1), \dots, \mathbf{A}_r(\lambda_r)$ where $\mathbf{A}_m(\lambda_m)$ is a matrix with the m th row and m th column being λ_m and λ_m^\top , 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 $\gamma(\theta) = \{\gamma_1(\theta), \dots, \gamma_r(\theta)\} \in \Gamma$ and $\lambda(\theta) = \{\lambda_1(\theta), \dots, \lambda_r(\theta)\} \in \Lambda$. We select $K_* = \lfloor c_p(n_*/\log p)^{q/2} \rfloor$ and define a collection $\mathcal{M}(\alpha, \nu)$ of covariance matrices as

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

where $\alpha > 0$ and $\nu > 0$ are two constants. Notice that each $\Sigma \in \mathcal{M}(\alpha, \nu)$ has value α along the main diagonal, and contains an $r \times r$ submatrix, say A , at the upper right corner, A^\top at the lower left corner and zero elsewhere. Write $\Sigma(\theta) = \{\sigma_{i,j}(\theta)\}_{p \times p}$. It holds that $\max_{\theta \in \Theta} \max_{i \in [p]} \sigma_{i,i}(\theta) = \alpha$ and $\max_{\theta \in \Theta} \max_{i \in [p]} \sum_{j=1}^p |\sigma_{i,j}(\theta)|^q \leq \alpha^q + c_p \nu^{q/2}$. 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 (7). Without loss of generality, we assume $\alpha = 1$ in the sequel and write $\mathcal{M}(1, \nu)$ as \mathcal{M} for simplification.

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

$$\begin{aligned} f_\theta &= \prod_{k=1}^{n_*} \frac{1}{(2\pi)^{p/2} |\Sigma(\theta)|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{u}_k^\top \Sigma^{-1}(\theta) \mathbf{u}_k\right\} \times \prod_{k=n_*+1}^n \frac{1}{\sqrt{2\pi} \sigma_{k,k}(\theta)} \exp\left\{-\frac{u_{k,k}^2}{2\sigma_{k,k}(\theta)}\right\} \\ &= \prod_{k=1}^{n_*} \frac{1}{(2\pi)^{p/2} |\Sigma(\theta)|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{u}_k^\top \Sigma^{-1}(\theta) \mathbf{u}_k\right\} \times \prod_{k=n_*+1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_{k,k}^2}{2}\right) \end{aligned}$$

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

$$\inf_{\Sigma} \max_{\theta \in \Theta} \mathbb{E}_{\mathcal{Z}|\theta} \left\{ \|\widehat{\Sigma} - \Sigma(\theta)\|_2^2 \right\} \geq \min_{(\theta, \theta') : H\{\gamma(\theta), \gamma(\theta')\} \geq 1} \frac{\|\Sigma(\theta) - \Sigma(\theta')\|_2^2}{H\{\gamma(\theta), \gamma(\theta')\}} \cdot \frac{r}{8} \cdot \min_{i \in [r]} \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\|,$$

where $H(\cdot, \cdot)$ is the Hamming distance, and $\bar{\mathbb{P}}_{i,a} = 2^{-(r-1)} |\Lambda|^{-1} \sum_{\theta \in \{\theta \in \Theta : \gamma_i(\theta) = a\}} \mathbb{P}_\theta$ for each $a \in \{0, 1\}$. In the sequel, we will show the following two results:

$$\min_{(\theta, \theta') : H\{\gamma(\theta), \gamma(\theta')\} \geq 1} \frac{\|\Sigma(\theta) - \Sigma(\theta')\|_2^2}{H\{\gamma(\theta), \gamma(\theta')\}} \gtrsim \frac{c_p^2}{p} \left(\frac{\log p}{n_*} \right)^{1-q} \tag{44}$$

and

$$\min_{i \in [r]} \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\| \gtrsim 1. \tag{45}$$

Recall $r = \lfloor p/2 \rfloor$. Then we will have [Theorem 4](#). The proofs of (44) and (45) are identical to that for Lemmas 5 and 6 in [Cai and Zhou \(2012b\)](#), respectively. Hence, we omit here. \square

8.8. Proof of [Theorem 5](#)

As shown in (23), $\mathbb{I}(i, j) = \mathbb{I}_1(i, j) + \mathbb{I}_2(i, j) + \mathbb{I}_3(i, j) + \mathbb{I}_4(i, j)$. Notice that Δ_K is a fixed integer. By Jensen's inequality, $\mathbb{E}\left\{\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds\right)^2\right\} \leq |t_{i,j,\ell} - t_{i,j,k}| \int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \mathbb{E}(\mu_{i,s}^2) ds \leq |t_{i,j,\ell} - t_{i,j,k}|^2 \max_{0 \leq s \leq T} \mathbb{E}(\mu_{i,s}^2)$, which implies

$$\max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{\ell: K \leq |\ell - k| \leq K + \Delta_K} \mathbb{E}\left\{\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds\right)^2\right\} \lesssim K^2 n_*^{-2}.$$

Due to $\mathbb{E}\left\{\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s}\right)^2\right\} = \mathbb{E}\left(\int_{t_{i,j,\ell} \wedge t_{i,j,k}}^{t_{i,j,\ell} \vee t_{i,j,k}} \sigma_{j,s}^2 ds\right) \leq |t_{i,j,\ell} - t_{i,j,k}| \max_{0 \leq s \leq T} \mathbb{E}(\sigma_{j,s}^2)$, we have

$$\max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{\ell: K \leq |\ell - k| \leq K + \Delta_K} \mathbb{E}\left\{\left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s}\right)^2\right\} \lesssim K n_*^{-1}.$$

By Cauchy–Schwarz inequality, $\max_{i,j \in [p]} |\mathbb{E}\{\mathbb{I}_1(i, j)\}| \lesssim K^2 n_*^{-2}$, $\max_{i,j \in [p]} |\mathbb{E}\{\mathbb{I}_3(i, j)\}| \lesssim K^{3/2} n_*^{-3/2}$ and $\max_{i,j \in [p]} |\mathbb{E}\{\mathbb{I}_4(i, j)\}| \lesssim K^{3/2} n_*^{-3/2}$. Notice that

$$\begin{aligned} \mathbb{E}\{\mathbb{I}_2(i, j)\} &= \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{K \leq |\ell - k| \leq K + \Delta_K} \mathbb{E}\left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds\right) \\ &= \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \left[\sum_{k=\ell-K-\Delta_K+1}^{\ell-K} \frac{\min\{k+K+\Delta_K, n_{i,j}\} - \ell}{N_{i,j,k}} \right. \\ &\quad + \sum_{k=\ell-K+1}^{\ell} \frac{|\min\{k+K+\Delta_K, n_{i,j}\} - k - K + 1|_+}{N_{i,j,k}} \\ &\quad + \sum_{k=\ell+1}^{\ell+K} \frac{|k - K + 1 - \max\{k - K - \Delta_K, 1\}|_+}{N_{i,j,k}} \\ &\quad \left. + \sum_{k=\ell+K+1}^{\ell+K+\Delta_K} \frac{\ell + 1 - \max\{k - K - \Delta_K, 1\}}{N_{i,j,k}} \right] \mathbb{E}\left(\int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds\right) \\ &:= \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} Q_{i,j,\ell} \cdot \mathbb{E}\left(\int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds\right), \end{aligned}$$

where we adopt the convention $N_{i,j,k} = \infty$ if $k > n_{i,j}$ or $k < 0$. For sufficiently large n , $N_{i,j,k}$ follows the formula (21). Since $K = o(n_*)$ and Δ_K is a fixed integer, for sufficiently large n , we have

$$Q_{i,j,\ell} \begin{cases} \lesssim K, & \text{if } 1 \leq \ell \leq 2(K + \Delta_K) - 1, \\ \frac{2K + \Delta_K}{2}, & \text{if } 2(K + \Delta_K) \leq \ell \leq n_{i,j} - 2(K + \Delta_K), \\ \lesssim K, & \text{if } n_{i,j} - 2(K + \Delta_K) + 1 \leq \ell \leq n_{i,j} - 1, \end{cases}$$

which implies that

$$\begin{aligned} \mathbb{E}\{\mathbb{I}_2(i, j)\} &= \frac{2K + \Delta_K}{4n_{i,j}} \sum_{\ell=2(K+\Delta_K)}^{n_{i,j}-2(K+\Delta_K)} \mathbb{E}\left(\int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds\right) + R_1(i, j) \\ &= \frac{2K + \Delta_K}{4n_{i,j}} \mathbb{E}\left(\int_{t_{i,j,1}}^{t_{i,j,n_{i,j}}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds\right) + R_2(i, j), \end{aligned}$$

where $\max_{i,j \in [p]} |R_1(i, j)| = O(K^2 n_*^{-2})$ and $\max_{i,j \in [p]} |R_2(i, j)| = O(K^2 n_*^{-2})$. Therefore, it follows from (23) that

$$\max_{i,j \in [p]} \left| \mathbb{E}\left\{\mathbb{I}(i, j) - \frac{2K + \Delta_K}{4n_{i,j}} \int_{t_{i,j,1}}^{t_{i,j,n_{i,j}}} \sigma_{i,s} \sigma_{j,s} \rho_{i,j,s} ds\right\} \right| \lesssim \frac{K^{3/2}}{n_*^{3/2}}.$$

We complete the proof of [Theorem 5](#). \square

8.9. Proof of Theorem 6

Write $\aleph = (n_*^{-1}K \log p)^{1/2}$. For Part (i), Theorem 1 implies $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| = O_p(\aleph)$. Due to $\max_{i,j \in [p]} |\hat{\psi}_{i,j}| = O_p(\log p)$ and $\aleph = o(1)$, by Triangle inequality, $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j}^{bc} - \sigma_{u,i,j}| \leq \max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| + \max_{i,j \in [p]} (4n_{i,j})^{-1}(2K + \Delta_K)|\hat{\psi}_{i,j}| = O_p(\aleph)$. Then Part (i) holds. To prove Part (ii), we define an event $\mathcal{E} = \{\max_{i,j \in [p]} (4n_{i,j})^{-1}(2K + \Delta_K)|\hat{\psi}_{i,j}| \leq \epsilon \aleph\}$ for some constant $\epsilon > 0$. Since $\max_{i,j \in [p]} |\hat{\psi}_{i,j}| = O_p(\log p)$ and $\aleph = o(1)$, then $\max_{i,j \in [p]} (4n_{i,j})^{-1}(2K + \Delta_K)|\hat{\psi}_{i,j}| = O_p(\aleph^2)$, which implies $\mathbb{P}(\mathcal{E}^c) = o(1)$. For any $C > 0$, by Markov's inequality, we have

$$\begin{aligned} \mathbb{P}(\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2 \geq Cc_p \aleph^{1-q}) &\leq \mathbb{P}(\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2 \geq Cc_p \aleph^{1-q}, \mathcal{E}) + \mathbb{P}(\mathcal{E}^c) \\ &\leq C^{-2} c_p^{-2} \aleph^{2(q-1)} \mathbb{E}\{\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2^2 I(\mathcal{E})\} + \mathbb{P}(\mathcal{E}^c). \end{aligned}$$

In the sequel, we will show $\mathbb{E}\{\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2^2 I(\mathcal{E})\} \lesssim c_p^2 \aleph^{2(1-q)}$. Based on this result, we know Part (ii) holds.

For each $i, j \in [p]$, we define the event $A_{i,j} = \{|\hat{\sigma}_{u,i,j}^{bc, \text{thre}} - \sigma_{u,i,j}| \leq 4 \min(|\sigma_{u,i,j}|, \alpha \aleph)\}$ with some constant $\alpha > 0$, and $d_{i,j} = (\hat{\sigma}_{u,i,j}^{bc, \text{thre}} - \sigma_{u,i,j})I(A_{i,j}^c)$. Write $\mathbf{D} = (d_{i,j})_{p \times p}$. Identical to (39) and the arguments below it, $\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2^2 \lesssim (\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|)^2 + \{\max_{i \in [p]} \sum_{j=1}^p |\hat{\sigma}_{u,i,j}^{bc, \text{thre}} - \sigma_{u,i,j}| I(A_{i,j}^c)\}^2 \lesssim (\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|)^2 + c_p^2 \aleph^{2(1-q)}$, which implies $\mathbb{E}\{\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2^2 I(\mathcal{E})\} \lesssim \mathbb{E}\{(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|)^2 I(\mathcal{E})\} + c_p^2 \aleph^{2(1-q)}$. Identical to (41), we have

$$\begin{aligned} \mathbb{E}\left\{\left(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|\right)^2 I(\mathcal{E})\right\} &\leq \underbrace{p \sum_{i,j=1}^p \mathbb{E}\{|\hat{\sigma}_{u,i,j}^{bc, \text{thre}} - \sigma_{u,i,j}|^2 I(\mathcal{E}) I[A_{i,j}^c \cap \{\hat{\sigma}_{u,i,j}^{bc, \text{thre}} = 0\}]\}}_I \\ &\quad + \underbrace{p \sum_{i,j=1}^p \mathbb{E}\{|\hat{\sigma}_{u,i,j}^{bc, \text{thre}} - \sigma_{u,i,j}|^2 I(\mathcal{E}) I[A_{i,j}^c \cap \{\hat{\sigma}_{u,i,j}^{bc, \text{thre}} = \hat{\sigma}_{u,i,j}^{bc}\}]\}}_II. \end{aligned}$$

Recall $\hat{\sigma}_{u,i,j}^{bc, \text{thre}} = \hat{\sigma}_{u,i,j}^{bc} I(|\hat{\sigma}_{u,i,j}^{bc}| \geq \beta \aleph)$ and $\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j}^{bc} - \sigma_{u,i,j}| \leq \epsilon \aleph$ restricted on \mathcal{E} . Then $I = p \sum_{i,j=1}^p \sigma_{u,i,j}^2 \mathbb{P}(|\sigma_{u,i,j}| \geq 4\alpha \aleph, |\hat{\sigma}_{u,i,j}^{bc}| < \beta \aleph, \mathcal{E}) \leq p \sum_{i,j=1}^p \sigma_{u,i,j}^2 \mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| \geq (4\alpha - \beta - \epsilon) \aleph\}$. Notice that $\max_{i,j \in [p]} \mathbb{E}\{|\hat{\sigma}_{u,i,j}^{bc} - \sigma_{u,i,j}|^4 I(\mathcal{E})\} \lesssim \max_{i,j \in [p]} \mathbb{E}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}|^4\} + \aleph^4 \lesssim 1$. Identical to (43), we have that $II \lesssim p \sum_{i,j=1}^p \{\mathbb{P}(|\hat{\sigma}_{u,i,j}^{bc} - \sigma_{u,i,j}| > 4\alpha \aleph, \mathcal{E})\}^{1/2} + p \sum_{i,j=1}^p \{\mathbb{P}\{|\hat{\sigma}_{u,i,j}^{bc} - \sigma_{u,i,j}| > (\beta - \alpha) \aleph, \mathcal{E}\}\}^{1/2} \leq p \sum_{i,j=1}^p \{\mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > (4\alpha - \epsilon) \aleph, \mathcal{E}\}\}^{1/2} + p \sum_{i,j=1}^p \{\mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > (\beta - \alpha - \epsilon) \aleph, \mathcal{E}\}\}^{1/2}$. Selecting $\alpha = 2\epsilon = \beta/2$ for some sufficiently large $\beta > 0$, applying the same arguments for bounding I and II there in Section 8.6, we have $I + II \lesssim o(\aleph^2)$, which implies $\mathbb{E}\{\|\hat{\Sigma}_u^{bc, \text{thre}} - \Sigma_u\|_2^2 I(\mathcal{E})\} \lesssim c_p^2 \aleph^{2(1-q)}$. We complete the proof of Theorem 6. \square

8.10. Proof of Part (i) in Theorem 7

For any $k \in [n_{i,j}]$, let $S_{i,j,k} = \{t_{i,j,\ell} : K \leq |\ell - k| \leq K + \Delta_K\}$. For any $i, j \in [p]$, we have that

$$\begin{aligned} \hat{\sigma}_{u,i,j} - \sigma_{u,i,j} &= \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (U_{i,t_{i,j,\ell}} - U_{i,t_{i,j,k}})(U_{j,t_{i,j,\ell}} - U_{j,t_{i,j,k}}) - \sigma_{u,i,j}}_{I(i,j)} \\ &\quad + \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}})(X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}})}_{II'(i,j)} \\ &\quad + \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (X_{i,t_{i,j,\ell}} - X_{i,t_{i,j,k}})(U_{j,t_{i,j,\ell}} - U_{j,t_{i,j,k}})}_{III'(i,j)} \\ &\quad + \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (U_{i,t_{i,j,\ell}} - U_{i,t_{i,j,k}})(X_{j,t_{i,j,\ell}} - X_{j,t_{i,j,k}})}_{IV'(i,j)}. \end{aligned}$$

Based on Proposition 1, we can obtain the convergence rate of $\max_{i,j \in [p]} |\text{II}(i, j)|$. Notice that $dX_{i,t} = \mu_{i,t} dt + \sigma_{i,t} dB_{i,t} + J_{i,t} dM_{i,t}$. In comparison to $\text{II}(i, j)$ specified in (23), we have

$$\begin{aligned} & \text{II}'(i, j) - \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{i,s} dM_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{j,s} dM_{j,s} \right) \\ &= \text{II}(i, j) + \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{i,s} ds \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{j,s} dM_{j,s} \right)}_{\text{II}_5(i,j)} \\ & \quad + \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{j,s} dM_{j,s} \right)}_{\text{II}_6(i,j)} \\ & \quad + \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{i,s} dM_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \sigma_{j,s} dB_{j,s} \right)}_{\text{II}_7(i,j)} \\ & \quad + \underbrace{\frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{i,s} dM_{i,s} \right) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} \mu_{j,s} ds \right)}_{\text{II}_8(i,j)}. \end{aligned} \tag{46}$$

For $\text{III}(i, j)$ specified in Section 8.1, we have $\text{III}'(i, j) = \text{III}(i, j) + (2n_{i,j})^{-1} \sum_{k=1}^{n_{i,j}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} (N_{i,j,\ell}^{-1} + N_{i,j,k}^{-1}) \left(\int_{t_{i,j,k}}^{t_{i,j,\ell}} J_{i,s} dM_{i,s} \right) U_{j,t_{i,j,k}}$. Propositions 2 and 3 give the convergence rates of $\max_{i,j \in [p]} |\text{II}(i, j)|$ and $\max_{i,j \in [p]} |\text{III}(i, j)|$, respectively. Write $\text{II}''(i, j) = \text{II}'(i, j) - \text{II}(i, j) - \varpi_{i,j}$ and $\text{III}''(i, j) = \text{III}'(i, j) - \text{III}(i, j)$. To prove Theorem 7, we need the following two propositions whose proofs are given in Sections 8.11 and 8.12, respectively.

Proposition 4. Under Assumptions 2 and 4–7, if $K(\log n_*)^{(2\iota+2\gamma+\iota\gamma)/(\iota\gamma)} = o(n_*)$, we have that

$$\max_{i,j \in [p]} \mathbb{P}\{|\text{II}''(i, j)| > v\} \lesssim \exp(-Cn_*K^{-1}v) + \exp\{-C(n_*K^{-1})^{\gamma/(2\iota+2\gamma+\iota\gamma)}\}$$

for any $v \gg n_*^{-1}K\lambda_*$, where γ and ι are specified in Assumptions 5 and 7, respectively. Furthermore, it holds that $\max_{i,j \in [p]} \mathbb{E}\{|\text{II}''(i, j)|^m\} \lesssim n_*^{(1-m)/2} K^{m/2}$ for any fixed positive integer m provided that $n_*^{-1}K\lambda_* = o(1)$.

Proposition 5. Under Assumptions 2–7, if $K(\log n_*)^{2+2/\iota} = o(n_*)$, we have that

$$\max_{i,j \in [p]} \mathbb{P}\{|\text{III}''(i, j)| \geq v\} \lesssim \exp\{-C(n_*K^{-1})^{(3\iota+2)/(4\iota+4)}v\} + \exp\{-C(n_*K^{-1})^{\gamma/(2\iota+2)}\}$$

for any $v \gg (n_*^{-1}K)^{(3\iota+2)/(4\iota+4)}\lambda_*$, where ι is specified in Assumption 7. Furthermore, it holds that $\max_{i,j \in [p]} \mathbb{E}\{|\text{III}''(i, j)|^m\} \lesssim n_*^{1/2}$ for any fixed positive integer m provided that $n_*^{-1}K\lambda_* = o(1)$.

Recall that $\hat{\sigma}_{u,i,j}^{\text{jump}} = \hat{\sigma}_{u,i,j} - \varpi_{i,j} = \sigma_{u,i,j} + \text{I}(i, j) + \text{II}(i, j) + \text{II}''(i, j) + \text{III}'(i, j) + \text{IV}'(i, j)$. Write $\aleph = (n_*^{-1}K \log p)^{1/2}$. We first consider the case with $\varphi < \infty$. Notice that $(1 + cx^{-1})^{-x} \geq e^{-c}$ for any $x > 0$ and $c > 0$, and $\text{III}'(i, j) = \text{III}(i, j) + \text{III}''(i, j)$. Since the tail probability of $\max_{i,j \in [p]} |\text{IV}'(i, j)|$ is the same as that of $\max_{i,j \in [p]} |\text{III}'(i, j)|$, by Propositions 1–5, if $K \gtrsim L_n$ and $K(\log n_*)^{\chi_1} = o(n_*)$ with $\chi_1 = \max\{(2\iota + 2\gamma + \iota\gamma)/(\iota\gamma), 2 + 2/\iota, 1 + 2/\gamma\}$, we have

$$\begin{aligned} \max_{i,j \in [p]} \mathbb{P}\{|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > v\} & \lesssim \{1 + n_*(K + L_n)^{-1}\rho^{-1}v^2\}^{-\rho/2} + v^{-1} \exp\{-C(n_*L_n^{-1}\rho^{-1}v)^{\varphi/(\varphi+1)}\} \\ & \quad + v^{-1} \exp(-Cn_*K^{-1}\rho^{-1}v) + \exp\{-C(n_*L_n^{-1}v)^{\varphi/(2\varphi+1)}\} \\ & \quad + \exp\{-C(n_*K^{-1})^{(3\iota+2)/(4\iota+4)}v\} + \exp\{-C(n_*K^{-1})^\eta\} \end{aligned} \tag{47}$$

for any $\rho \geq 1$ and $(n_*^{-1}K)^{2/(\gamma+4)} \gg v \gg \max\{\exp(-CL_n^{-\varphi}K^\varphi), (n_*^{-1}K)^{(3\iota+2)/(4\iota+4)}\lambda_*\}$, where $\eta = \min\{\gamma/(\gamma + 4), \iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2)\}$. Since $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, $\lambda_*^2(n_*^{-1}K)^{\gamma/(2\iota+2)}(\log p)^{-1} = o(1)$ and $\log p = o\{(n_*K^{-1})^{\gamma/(\gamma+4)}\}$, then $(n_*^{-1}K)^{2/(\gamma+4)} \gg \aleph \gg \max\{\exp(-CL_n^{-\varphi}K^\varphi), (n_*^{-1}K)^{(3\iota+2)/(4\iota+4)}\lambda_*\}$. Given a sufficiently large constant $\alpha > 0$, we have $\mathbb{E}\{|\hat{\Sigma}_u^{\text{jump}} - \Sigma_u|\} \leq \mathbb{E}\{\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| I(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| \leq \alpha\aleph)\} + \mathbb{E}\{\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| I(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > \alpha\aleph)\} =: A_1^* + A_2^*$. It is easy to see that $A_1^* \leq \alpha\aleph$. By Cauchy–Schwarz inequality, we have $A_2^* \leq \sum_{i,j=1}^p \mathbb{E}\{|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| I(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > \alpha\aleph)\} \leq p^2 \max_{i,j \in [p]} \{\mathbb{E}\{|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}|^2\}\}^{1/2} \cdot \max_{i,j \in [p]} \{\mathbb{P}\{|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > \alpha\aleph\}\}^{1/2}$. Let $\rho \asymp \log p \geq 1$.

Since $K \log p = o(n_*)$, (47) implies $\max_{i,j \in [p]} \mathbb{P}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > \alpha \aleph) \lesssim p^{-2w} + \exp[-C\{n_* K L_n^{-2} (\log p)^{-1}\}^{\varphi/(2\varphi+2)}] + \exp[-C\{n_* K^{-1} (\log p)^{-1}\}^{1/2}] + \exp\{-C(n_* K^{-1})^{\nu/(4\iota+4)} (\log p)^{1/2}\} + \exp\{-C(n_* K^{-1})^\eta\}$ with some sufficiently large $w > 0$, where $w \rightarrow \infty$ as $\alpha \rightarrow \infty$. Due to $\max_{i,j \in [p]} \mathbb{E}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}|^2) \lesssim n_*^{1/2} \leq p^{2c_*}$ with $c_* = 1/(4\kappa)$, where κ is specified in the beginning of Section 3, if $\log p = o\{\min\{(n_* L_n^{-2} K)^{\varphi/(3\varphi+2)}, (n_* K^{-1})^{\chi_2}\}\}$ with $\chi_2 = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2), \gamma/(\gamma + 4), 1/3\}$, then $A_*^2 \lesssim p^{2+c_*-w} + \exp[-C\{n_* K L_n^{-2} (\log p)^{-1}\}^{\varphi/(2\varphi+2)}] + \exp[-C\{n_* K^{-1} (\log p)^{-1}\}^{1/2}] + \exp\{-C(n_* K L_n^{-2} \log p)^{\varphi/(4\varphi+2)}\} + \exp\{-C(n_* K^{-1})^{\nu/(4\iota+4)} (\log p)^{1/2}\} + \exp\{-C(n_* K^{-1})^\eta\} = o\{(n_*^{-1} K \log p)^4\}$. Hence, $\sup_{\mathcal{P}_3} \mathbb{E}(|\hat{\Sigma}_u^{\text{jump}} - \Sigma_u|_\infty) \lesssim (n_*^{-1} K \log p)^{1/2}$ provided that $K^{-\varphi} L_n^\varphi \log\{n_* (K \log p)^{-1}\} = o(1)$, $\lambda_*^2 (n_*^{-1} K)^{\nu/(2\iota+2)} (\log p)^{-1} = o(1)$, $K \gtrsim L_n$ and $K(\log n_*)^{\chi_1} = o(n_*)$ with $\chi_1 = \max\{(2\iota + 2\gamma + \iota\gamma)/(\iota\gamma), 2 + 2/\iota, 1 + 2/\gamma\}$, and $\log p = o\{\min\{(n_* L_n^{-2} K)^{\varphi/(3\varphi+2)}, (n_* K^{-1})^{\chi_2}\}\}$ with $\chi_2 = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2), \gamma/(\gamma + 4), 1/3\}$.

Now we consider the case with $\varphi = \infty$. As we discussed in Remark 5(i), if $\{\mathbf{U}_k\}$ is an independent sequence, we can select $L_n = 1/2$. Due to $K \geq 1$, we have $K > L_n$ in this case. Without loss of generality, we can always assume $K > L_n$ when $\varphi = \infty$. Based on Remark 5, it holds that $\{1 + n_*(K + L_n)^{-1} \rho^{-1} v^2\}^{-\rho/2} + v^{-1} \exp(-C n_* K^{-1} \rho^{-1} v)$ for any $v > 0$ and $\rho \geq 1$ under either of the scenarios: (i) $\{\mathbf{U}_k\}$ is an independent sequence, and (ii) $\{\mathbf{U}_k\}$ is an L_n -dependent sequence. Repeating the arguments for $\varphi < \infty$, we have $\sup_{\mathcal{P}_3} \mathbb{E}(|\hat{\Sigma}_u^{\text{jump}} - \Sigma_u|_\infty) \lesssim (n_*^{-1} K \log p)^{1/2}$ provided that $\lambda_*^2 (n_*^{-1} K)^{\nu/(2\iota+2)} (\log p)^{-1} = o(1)$, $K(\log n_*)^{\chi_1} = o(n_*)$ with $\chi_1 = \max\{(2\iota + 2\gamma + \iota\gamma)/(\iota\gamma), 2 + 2/\iota, 1 + 2/\gamma\}$ and $\log p = o\{(n_* K^{-1})^{\chi_2}\}$ with $\chi_2 = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2), \gamma/(\gamma + 4), 1/3\}$. We complete the proof of part (i) in Theorem 7. \square

8.11. Proof of Proposition 4

Recall that $\xi = \max_{i,j \in [p]} \max_{k \in [n_i]} \max_{t_{i,j,\ell} \in S_{i,j,k}} |t_{i,j,\ell} - t_{i,j,k}| \asymp n_*^{-1} K$, $S_{i,j,k} = \{t_{i,j,\ell} : K \leq |\ell - k| \leq K + \Delta_K\}$ and $N_{i,j,k} = |S_{i,j,k}|$. In the sequel, we will bound the tail probabilities of $\max_{i,j \in [p]} |\Pi_5(i, j)|$, $\max_{i,j \in [p]} |\Pi_6(i, j)|$, $\max_{i,j \in [p]} |\Pi_7(i, j)|$ and $\max_{i,j \in [p]} |\Pi_8(i, j)|$, respectively.

We first bound the tail probabilities of $\max_{i,j \in [p]} |\Pi_5(i, j)|$ and $\max_{i,j \in [p]} |\Pi_8(i, j)|$. Notice that for sufficiently large n , $\min_{k \in [n_i]} N_{i,j,k} = \Delta_K + 1$ and $\max_{k \in [n_i]} N_{i,j,k} = 2\Delta_K + 2$. Since Δ_K is a fixed constant, it holds that

$$\begin{aligned} |\Pi_5(i, j)| &\leq \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \frac{1}{N_{i,j,k}} \sum_{K \leq \ell - k \leq K + \Delta_K} \left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |\mu_{i,s}| ds \right) \left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |J_{j,s}| dM_{j,s} \right) \\ &\leq \left(\max_{k \in [n_i]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |\mu_{i,s}| ds \right) \\ &\quad \times \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \left[\sum_{k=\ell-K-\Delta_K+1}^{\ell-K} \frac{\min\{k + K + \Delta_K, n_{i,j}\} - \ell}{N_{i,j,k}} \right. \\ &\quad \left. + \sum_{k=\ell-K+1}^{\ell} \frac{|\min\{k + K + \Delta_K, n_{i,j}\} - k - K + 1|_+}{N_{i,j,k}} \right. \\ &\quad \left. + \sum_{k=\ell+1}^{\ell+K} \frac{|k - K + 1 - \max\{k - K - \Delta_K, 1\}|_+}{N_{i,j,k}} \right. \\ &\quad \left. + \sum_{k=\ell+K+1}^{\ell+K+\Delta_K} \frac{\ell + 1 - \max\{k - K - \Delta_K, 1\}}{N_{i,j,k}} \right] \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |J_{j,s}| dM_{j,s} \\ &\lesssim \left(\max_{k \in [n_i]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} K |\mu_{i,s}| ds \right) \times \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |J_{j,s}| dM_{j,s} \\ &:= Q_{i,j}^\mu \cdot \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |J_{j,s}| dM_{j,s}, \end{aligned} \tag{48}$$

where we adopt the convention $N_{i,j,k} = \infty$ if $k > n_{i,j}$ or $k < 0$. For any constant $d_1 \in (0, K^{-1} \xi^{-1})$, define $\mathcal{E}_{j,d_1} = \{\sup_{0 \leq s \leq T} |J_{j,s}| > d_1\}$. Recall $(\Delta M_{i,\cdot})_{i,j} = M_{i,t_{i,j},n_{i,j}} - M_{i,t_{i,j},1}$ for any $i, j \in [p]$. By Assumption 7, for any $v > 0$, we have

$$\begin{aligned} \mathbb{P}\left(Q_{i,j}^\mu \cdot \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |J_{j,s}| dM_{j,s} \geq v\right) &\leq \mathbb{P}\left(Q_{i,j}^\mu \cdot \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |J_{j,s}| dM_{j,s} \geq v, \mathcal{E}_{j,d_1}^c\right) + \mathbb{P}(\mathcal{E}_{j,d_1}) \\ &\leq \mathbb{P}\{d_1 Q_{i,j}^\mu \cdot (\Delta M_{j,\cdot})_{j,i} \geq 2n_{i,j} v\} + C \exp(-Cd_1). \end{aligned}$$

It follows from Bonferroni inequality that $\mathbb{P}(d_1 Q_{ij}^\mu \geq v) \lesssim n_* \max_{k \in [n_{ij}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{P}(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} K d_1 |\mu_{i,s}| ds \geq v)$ for any $v > 0$. For any $\theta \in (0, C_5 d_1^{-1} K^{-1} \xi^{-1}]$, by Jensen's inequality, **Assumption 4** implies that $\mathbb{E}\{\exp(\theta \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} K d_1 |\mu_{i,s}| ds)\} \leq |t_{i,j,\ell} - t_{i,j,k}|^{-1} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \mathbb{E}\{\exp(\theta K d_1 |t_{i,j,\ell} - t_{i,j,k}| |\mu_{i,s}|)\} ds \leq \sup_{0 \leq s \leq T} \mathbb{E}\{\exp(\theta K d_1 \xi |\mu_{i,s}|)\} \lesssim \exp(CK^2 d_1^2 \xi^2 \theta^2)$. Selecting $\theta \asymp K^{-1} d_1^{-1} \xi^{-1}$, by Markov's inequality, $\max_{k \in [n_{ij}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{P}(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} K d_1 |\mu_{i,s}| ds \geq v) \lesssim \exp(-CK^{-1} d_1^{-1} \xi^{-1} v)$ for any $v > 0$, which implies that

$$\max_{i,j \in [p]} \mathbb{P}(d_1 Q_{ij}^\mu \geq v) \lesssim n_* \exp(-CK^{-1} d_1^{-1} \xi^{-1} v) \tag{49}$$

for any $v > 0$. By **Assumption 6**, applying Proposition 2.9 of **Wainwright (2019)**, we have $\mathbb{P}[(\Delta M_{j,\cdot})_{j,i} - \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}\}] \geq v] \leq 2 \exp[-v^2 / \{4\lambda_j(t_{j,i,n_{j,i}} - t_{j,i,1})\}] + 2 \exp\{-v / (2C_{10})\}$ for any $v > 0$. By **Assumption 2** and (4), $(t_{j,i,n_{j,i}} - t_{j,i,1}) \asymp 1$ and $n_{i,j} \asymp n_*$, which implies that $\mathbb{P}[(\Delta M_{j,\cdot})_{j,i} - \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}\}] \geq n_{i,j} v] \lesssim \exp(-Cn_*^2 v^2 / \lambda_j) + \exp(-Cn_*)$ for any $v > 0$. By **Assumption 6(ii)**, we have $\max_{i,j \in [p]} |n_{i,j}^{-1} \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}\}| \lesssim n_*^{-1} \lambda_*$. Then for any $v \gg n_*^{-1} \lambda_*$, it holds that

$$\begin{aligned} \max_{i,j \in [p]} \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i} \geq n_{i,j} v\} &\leq \max_{i,j \in [p]} \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i} - \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}\} \geq n_{i,j} v - |\mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}\}|\} \\ &\lesssim \max_{i,j \in [p]} \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i} - \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}\} \geq n_{i,j} v / 2\} \\ &\lesssim \exp(-C\lambda_*^{-1} n_*^2 v^2) + \exp(-Cn_*) \lesssim \exp(-Cn_* v). \end{aligned} \tag{50}$$

Combining (49) and (50), we have that $\mathbb{P}\{d_1 Q_{ij}^\mu (\Delta M_{j,\cdot})_{j,i} \geq 2n_{i,j} v\} \leq \mathbb{P}\{d_1 Q_{ij}^\mu \geq K\} + \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i} \geq 2K^{-1} n_{i,j} v\} \lesssim n_* \exp(-Cd_1^{-1} \xi^{-1}) + \exp(-Cn_* K^{-1} v)$ for any $v \gg n_*^{-1} K \lambda_*$. Then $\mathbb{P}\{Q_{ij}^\mu \cdot (2n_{i,j})^{-1} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{j,s}| dM_{j,s} \geq v\} \lesssim n_* \exp(-Cd_1^{-1} \xi^{-1}) + \exp(-Cn_* K^{-1} v) + \exp(-Cd_1)$ for any $v \gg n_*^{-1} K \lambda_*$. Notice that $\xi \asymp n_*^{-1} K$. Thus, by (48), $\max_{i,j \in [p]} \mathbb{P}\{\|II_5(i, j)\| \geq v\} \lesssim \exp(-Cn_* K^{-1} v) + n_* \exp(-Cd_1^{-1} n_* K^{-1}) + \exp(-Cd_1)$ for any $v \gg n_*^{-1} K \lambda_*$. In order to make $\max_{i,j \in [p]} \|II_5(i, j)\| = O_p(n_*^{-1/2} K^{1/2} \log^{1/2} p)$ and p diverge as fast as possible, we require $\log p = o(n_* K^{-1})$ and select $d_1 \asymp (n_* K^{-1})^{1/(1+\iota)}$. Then if $K(\log n_*)^{(1+\iota)/\iota} = o(n_*)$, it holds that

$$\max_{i,j \in [p]} \mathbb{P}\{\|II_5(i, j)\| \geq v\} \lesssim \exp(-Cn_* K^{-1} v) + \exp\{-C(n_* K^{-1})^{\iota/(1+\iota)}\} \tag{51}$$

for any $v \gg n_*^{-1} K \lambda_*$. Identically, if $K(\log n_*)^{(1+\iota)/\iota} = o(n_*)$, we also have

$$\max_{i,j \in [p]} \mathbb{P}\{\|II_8(i, j)\| \geq v\} \lesssim \exp(-Cn_* K^{-1} v) + \exp\{-C(n_* K^{-1})^{\iota/(1+\iota)}\} \tag{52}$$

for any $v \gg n_*^{-1} K \lambda_*$.

Now we consider $II_6(i, j)$ and $II_7(i, j)$. Analogously to (48), we have

$$\begin{aligned} \|II_6(i, j)\| &\leq \frac{1}{2n_{ij}} \sum_{k=1}^{n_{ij}} \frac{1}{N_{i,j,k}} \sum_{K \leq \ell - k \leq K + \Delta K} \left| \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right| \left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |U_{j,s}| dM_{j,s} \right) \\ &\lesssim \left(\max_{k \in [n_{ij}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} K \left| \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right| \right) \times \frac{1}{2n_{ij}} \sum_{\ell=1}^{n_{ij}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{j,s}| dM_{j,s} \\ &:= Q_{ij}^\sigma \cdot \frac{1}{2n_{ij}} \sum_{\ell=1}^{n_{ij}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{j,s}| dM_{j,s}. \end{aligned} \tag{53}$$

For any constant $d_2 \in (0, K^{-1} \xi^{-1}]$, define the event $\mathcal{E}_{j,d_2} = \{\sup_{0 \leq s \leq T} |U_{j,s}| > d_2\}$. By **Assumption 7**,

$$\begin{aligned} \mathbb{P}\left(Q_{ij}^\sigma \cdot \frac{1}{2n_{ij}} \sum_{\ell=1}^{n_{ij}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{j,s}| dM_{j,s} \geq v\right) &\leq \mathbb{P}\left(Q_{ij}^\sigma \cdot \frac{1}{2n_{ij}} \sum_{\ell=1}^{n_{ij}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{j,s}| dM_{j,s} \geq v, \mathcal{E}_{j,d_2}^c\right) + \mathbb{P}(\mathcal{E}_{j,d_2}) \\ &\leq \mathbb{P}\{d_2 Q_{ij}^\sigma \cdot (\Delta M_{j,\cdot})_{j,i} \geq 2n_{i,j} v\} + C \exp(-Cd_2^2) \end{aligned}$$

for any $v > 0$. By Bonferroni inequality, we have

$$\begin{aligned} \mathbb{P}(d_2 Q_{ij}^\sigma \geq v) &\leq \sum_{k=1}^{n_{ij}} \sum_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{P}\left(K d_2 \left| \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right| \geq v\right) \\ &\lesssim n_* \max_{k \in [n_{ij}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{P}\left(d_2 K \left| \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right| \geq v\right) \end{aligned} \tag{54}$$

for any $v > 0$. For any constant $d \in (0, d_2^{-1/2}K^{-1/2}\xi^{-1/2}]$, define a stopping time $\Gamma_{i,d} = T \wedge \inf\{t > 0 : \sup_{0 \leq s \leq t} \sigma_{i,s} > d\}$. Then it holds that

$$\begin{aligned} & \mathbb{P}\left(d_2K \left| \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \right| \geq v, \Gamma_{i,d} = T\right) \\ & \lesssim \exp(-C\theta v^2) \mathbb{E}\left[\exp\left\{\theta d_2^2 K^2 \left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s}\right)^2\right\} I(\Gamma_{i,d} = T)\right] \end{aligned} \tag{55}$$

for any $\theta > 0$. Identical to (27) and (28), $\mathbb{E}[\exp\{\theta d_2^2 K^2 (\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s})^2\} I(\Gamma_{i,d} = T)] \lesssim 1$. With $\theta = d_2^{-2}d^{-2}K^{-2}\xi^{-1}/4$, (55) implies $\mathbb{P}(d_2K \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \geq v, \Gamma_{i,d} = T) \lesssim \exp(-Cd_2^{-2}d^{-2}K^{-2}\xi^{-1}v^2)$. By Assumption 5, $\max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{P}(d_2K \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s} \geq v) \lesssim \exp(-Cd_2^{-2}d^{-2}K^{-2}\xi^{-1}v^2) + \exp(-Cd^\gamma)$ for any $v > 0$. Then (54) implies that $\mathbb{P}(d_2Q_{i,j}^\sigma \geq v) \lesssim n_* \exp(-Cd_2^{-2}d^{-2}K^{-2}\xi^{-1}v^2) + n_* \exp(-Cd^\gamma)$ for any $v > 0$. Together with (50), we have $\mathbb{P}\{d_2Q_{i,j}^\sigma(\Delta M_{j,\cdot})_{j,i} \geq 2n_{i,j}v\} \leq \mathbb{P}(d_2Q_{i,j}^\sigma \geq K) + \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i} \geq 2K^{-1}n_{i,j}v\} \lesssim n_* \exp(-Cd_2^{-2}d^{-2}\xi^{-1}) + n_* \exp(-Cd^\gamma) + \exp(-Cn_*K^{-1}v)$ for any $v \gg n_*^{-1}K\lambda_*$, which implies $\max_{i,j \in [p]} \mathbb{P}\{\|II_6(i,j)\| \geq v\} \lesssim n_* \exp(-Cd_2^{-2}d^{-2}\xi^{-1}) + n_* \exp(-Cd^\gamma) + \exp(-Cd_2^2) + \exp(-Cn_*K^{-1}v)$ for any $v \gg n_*^{-1}K\lambda_*$. To make $\max_{i,j \in [p]} \|II_6(i,j)\| = O_p(n_*^{1/2}K^{1/2} \log^{1/2} p)$ and p diverge as fast as possible, we require $\log p = o(n_*K^{-1})$, and select $d_2 \asymp (n_*K^{-1})^{\gamma/(2\iota+2\gamma+\iota\gamma)}$ and $d \asymp (d_2^{-2}n_*K^{-1})^{1/(\gamma+2)}$. Then if $K(\log n_*)^{(2\iota+2\gamma+\iota\gamma)/(\iota\gamma)} = o(n_*)$, we have

$$\max_{i,j \in [p]} \mathbb{P}\{\|II_6(i,j)\| \geq v\} \lesssim \exp(-Cn_*K^{-1}v) + \exp\{-C(n_*K^{-1})^{\iota\gamma/(2\iota+2\gamma+\iota\gamma)}\} \tag{56}$$

for any $v \gg n_*^{-1}K\lambda_*$. Analogously, if $K(\log n_*)^{(2\iota+2\gamma+\iota\gamma)/(\iota\gamma)} = o(n_*)$, it holds that

$$\max_{i,j \in [p]} \mathbb{P}\{\|II_7(i,j)\| \geq v\} \lesssim \exp(-Cn_*K^{-1}v) + \exp\{-C(n_*K^{-1})^{\iota\gamma/(2\iota+2\gamma+\iota\gamma)}\} \tag{57}$$

for any $v \gg n_*^{-1}K\lambda_*$. Notice that $\iota\gamma/(2\iota+2\gamma+\iota\gamma) \leq \iota/(1+\iota)$. Combining (51), (52), (56) and (57), if $K(\log n_*)^{(2\iota+2\gamma+\iota\gamma)/(\iota\gamma)} = o(n_*)$, it holds that $\max_{i,j \in [p]} \mathbb{P}\{\|II''(i,j)\| > v\} \lesssim \exp(-Cn_*K^{-1}v) + \exp\{-C(n_*K^{-1})^{\iota\gamma/(2\iota+2\gamma+\iota\gamma)}\}$ for any $v \gg n_*^{-1}K\lambda_*$.

By (48) and Cauchy-Schwarz inequality, for any positive integer m , we have $\mathbb{E}\{\|II_5(i,j)\|^m\} \lesssim [\mathbb{E}\{(K^{-1}Q_{i,j}^\sigma)^{2m}\}]^{1/2} \cdot [\mathbb{E}\{(n_i^{-1}K \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |J_{j,s}| dM_{j,s})^{2m}\}]^{1/2} =: E_{i,j,1} \cdot E_{i,j,2}$. By Assumption 4 and Jensen's inequality, it holds that

$$\begin{aligned} E_{i,j,1}^2 & \lesssim n_* \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\left\{\left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |\mu_{i,s}| ds\right)^{2m}\right\} \\ & \lesssim n_* \xi^{2m-1} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sup_{0 \leq s \leq T} \mathbb{E}\{|\mu_{i,s}|^{2m}\} ds \lesssim n_* \xi^{2m}. \end{aligned}$$

If $\lambda_* n_*^{-1}K = o(1)$, it follows from Assumptions 6 and 7 that

$$\begin{aligned} E_{i,j,2}^2 & \leq \mathbb{E}\left[\left(\sup_{0 \leq s \leq T} |J_{i,s}|^{2m}\right) \{Kn_{i,j}^{-1}(\Delta M_{j,\cdot})_{j,i}\}^{2m}\right] \\ & \lesssim \left\{\mathbb{E}\left(\sup_{0 \leq s \leq T} |J_{i,s}|^{4m}\right)\right\}^{1/2} \cdot (\mathbb{E}\{[Kn_{i,j}^{-1}(\Delta M_{j,\cdot})_{j,i}]^{4m}\})^{1/2} \\ & \lesssim (\mathbb{E}\{\exp\{Kn_{i,j}^{-1}(\Delta M_{j,\cdot})_{j,i}\}\})^{1/2} \lesssim 1. \end{aligned} \tag{58}$$

Thus, $\max_{i,j \in [p]} \mathbb{E}\{\|II_5(i,j)\|^m\} \lesssim n_*^{1/2} \xi^m$. Analogously, we have $\max_{i,j \in [p]} \mathbb{E}\{\|II_8(i,j)\|^m\} \lesssim n_*^{1/2} \xi^m$. By (53) and Cauchy-Schwarz inequality, it holds that $\mathbb{E}\{\|II_6(i,j)\|^m\} \lesssim [\mathbb{E}\{(K^{-1}Q_{i,j}^\sigma)^{2m}\}]^{1/2} \cdot [\mathbb{E}\{(n_i^{-1}K \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |J_{j,s}| dM_{j,s})^{2m}\}]^{1/2} =: E_{i,j,3} \cdot E_{i,j,2}$ for any positive integer m . By Assumptions 4 and 5, Burkholder-Davis-Gundy inequality implies

$$E_{i,j,3}^2 \lesssim n_* \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\left\{\left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} \sigma_{i,s} dB_{i,s}\right)^{2m}\right\} \lesssim n_* \xi^m.$$

Together with (58), $\max_{i,j \in [p]} \mathbb{E}\{\|II_6(i,j)\|^m\} \lesssim n_*^{1/2} \xi^{m/2}$. Analogously, $\max_{i,j \in [p]} \mathbb{E}\{\|II_7(i,j)\|^m\} \lesssim n_*^{1/2} \xi^{m/2}$. Hence, we have $\max_{i,j \in [p]} \mathbb{E}\{\|II''(i,j)\|^m\} \lesssim n_*^{1/2} \xi^{m/2}$ for any positive integer m provided that $\lambda_* n_*^{-1}K = o(1)$. We complete the proof of Proposition 4. \square

8.12. Proof of Proposition 5

Notice that for sufficiently large n , $\min_{k \in [n_{i,j}]} N_{i,j,k} = \Delta_K + 1$ and $\max_{k \in [n_{i,j}]} N_{i,j,k} = 2\Delta_K + 2$. Analogously to (48), we have that

$$\begin{aligned} |\text{III}''(i, j)| &\lesssim \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \sum_{K \leq |\ell-k| \leq K + \Delta_K} \left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |U_{i,s}| dM_{i,s} \right) |U_{j,t_{i,j,k}}| \\ &\lesssim \left(K \max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}| \right) \times \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s}. \end{aligned} \tag{59}$$

For any constant $d_3 > 0$, define the event $\mathcal{E}_{i,d_3} = \{\sup_{0 \leq s \leq T} |U_{i,s}| > d_3\}$. By Assumption 7, it holds that

$$\begin{aligned} &\mathbb{P} \left\{ \left(K \max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}| \right) \cdot \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s} \geq v \right\} \\ &\leq \mathbb{P} \left\{ \left(K \max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}| \right) \cdot \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s} \geq v, \mathcal{E}_{i,d_3}^c \right\} + \mathbb{P}(\mathcal{E}_{i,d_3}) \\ &\leq \mathbb{P} \left\{ \left(d_3 K \max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}| \right) (\Delta M_{i,\cdot})_{i,j} \geq 2n_{i,j}v \right\} + C \exp(-Cd_3^\iota) \end{aligned} \tag{60}$$

for any $v > 0$. By Bonferroni inequality and Assumption 3, $\mathbb{P}(d_3 K \max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}| \geq v) \leq \sum_{k=1}^{n_{i,j}} \mathbb{P}(|U_{j,t_{i,j,k}}| \geq d_3^{-1}K^{-1}v) \lesssim n_* \max_{k \in [n_{i,j}]} \mathbb{P}(|U_{j,t_{i,j,k}}| \geq d_3^{-1}K^{-1}v) \lesssim n_* \exp(-Cd_3^{-2}K^{-2}v^2)$ for any $v > 0$. Recall $\xi \asymp n_*^{-1}K$. Together with (50), for any $\delta \in (0, 1)$, we have that $\mathbb{P}(\{d_3 K \max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}|\} (\Delta M_{i,\cdot})_{i,j} \geq 2n_{i,j}v) \leq \mathbb{P}(d_3 K \max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}| \geq K\xi^{-\delta}) + \mathbb{P}(\{(\Delta M_{i,\cdot})_{i,j} \geq 2n_{i,j}K^{-1}\xi^\delta v\}) \lesssim n_* \exp(-Cd_3^{-2}\xi^{-2\delta}) + \exp(-C\xi^{\delta-1}v)$ for any $v \gg \xi^{1-\delta}\lambda_*$. By (59) and (60), $\max_{i,j \in [p]} \mathbb{P}\{|\text{III}''(i, j)| \geq v\} \lesssim n_* \exp(-Cd_3^{-2}\xi^{-2\delta}) + \exp(-C\xi^{\delta-1}v) + \exp(-Cd_3^\iota)$ for any $v \gg \xi^{1-\delta}\lambda_*$. To make $\max_{i,j \in [p]} |\text{III}''(i, j)| = O_p(n_*^{-1/2}K^{1/2} \log^{1/2} p)$, it suffices to require $\log p = o(\min\{d_3^{-2}\xi^{-2\delta}, d_3^\iota, \xi^{2\delta-1}\})$, $\lambda_*^\iota \xi^{1-2\delta}(\log p)^{-1} = o(1)$ and $\log n_* = o(d_3^{-2}\xi^{-2\delta})$. To make p diverge as fast as possible, we select δ such that $d_3^{-2}\xi^{-2\delta} \asymp d_3^\iota \asymp \xi^{2\delta-1}$. Then $\delta = (\iota+2)/(4\iota+4)$ and $d_3 = \xi^{-1/(2\iota+2)}$. Hence, if $K(\log n_*)^{2(\iota+1)/\iota} = o(n_*)$, we have $\max_{i,j \in [p]} \mathbb{P}\{|\text{III}''(i, j)| \geq v\} \lesssim \exp\{-C(n_*K^{-1})^{(3\iota+2)/(4\iota+4)}v\} + \exp\{-C(n_*K^{-1})^{\iota/(2\iota+2)}\}$ for any $v \gg (n_*^{-1}K)^{(3\iota+2)/(4\iota+4)}\lambda_*$.

By (59) and Cauchy-Schwarz inequality, we have $\mathbb{E}\{|\text{III}''(i, j)|^m\} \lesssim \{\mathbb{E}(\max_{k \in [n_{i,j}]} |U_{j,t_{i,j,k}}|^{2m})\}^{1/2} \cdot [\mathbb{E}\{(n_*^{-1}K \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s})^{2m}\}]^{1/2} =: F_{i,j,1} \cdot F_{i,j,2}$ for any positive integer m . It follows Assumption 3 that $F_{i,j,1}^2 \lesssim n_* \max_{k \in [n_{i,j}]} \mathbb{E}(U_{j,t_{i,j,k}}^{2m}) \lesssim n_*$. Analogously to (58), we have $F_{i,j,2}^2 \lesssim 1$ if $\lambda_* n_*^{-1}K = o(1)$. Hence, $\max_{i,j \in [p]} \mathbb{E}\{|\text{III}''(i, j)|^m\} \lesssim n_*^{1/2}$ for any positive integer m provided that $\lambda_* n_*^{-1}K = o(1)$. We complete the proof of Proposition 5. \square

8.13. Proof of Part (ii) in Theorem 7

We first consider the case with $\varphi < \infty$. Write $\aleph = (n_*^{-1}K \log p)^{1/2}$. For each $i, j \in [p]$, we define the event $A_{i,j} = \{|\hat{\sigma}_{u,i,j}^{\text{jump,thre}} - \sigma_{u,i,j}| \leq 4 \min\{|\sigma_{u,i,j}|, \alpha \aleph\}\}$ with some constant $\alpha > 0$, and $d_{i,j} = (\hat{\sigma}_{u,i,j}^{\text{jump,thre}} - \sigma_{u,i,j})I(A_{i,j}^c)$. Identical to (40), we have $\mathbb{E}(\|\hat{\Sigma}_u^{\text{jump,thre}} - \Sigma_u\|_2^2) \lesssim \mathbb{E}\{(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|)^2\} + c_p^2 \aleph^{2(1-q)}$. It holds that

$$\begin{aligned} \mathbb{E} \left\{ \left(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}| \right)^2 \right\} &\leq p \sum_{i,j=1}^p \mathbb{E}\{|\hat{\sigma}_{u,i,j}^{\text{jump,thre}} - \sigma_{u,i,j}|^2 I(A_{i,j}^c)\} \\ &= p \underbrace{\sum_{i,j=1}^p \mathbb{E}\left(|\hat{\sigma}_{u,i,j}^{\text{jump,thre}} - \sigma_{u,i,j}|^2 I[A_{i,j}^c \cap \{\hat{\sigma}_{u,i,j}^{\text{jump,thre}} = 0\}]\right)}_I \\ &\quad + p \underbrace{\sum_{i,j=1}^p \mathbb{E}\left(|\hat{\sigma}_{u,i,j}^{\text{jump,thre}} - \sigma_{u,i,j}|^2 I[A_{i,j}^c \cap \{\hat{\sigma}_{u,i,j}^{\text{jump,thre}} = \hat{\sigma}_{u,i,j}^{\text{jump}}\}]\right)}_{II}. \end{aligned}$$

Recall $\hat{\sigma}_{u,i,j}^{\text{jump,thre}} = \hat{\sigma}_{u,i,j}^{\text{jump}} I(|\hat{\sigma}_{u,i,j}^{\text{jump}}| \geq \beta \aleph)$ for any $i, j \in [p]$. Identical to (42), $I \leq p \sum_{i,j=1}^p \sigma_{u,i,j}^2 \mathbb{P}\{|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| \geq (4\alpha - \beta)\aleph\}$. Selecting $\alpha = \beta/2$ and β being sufficiently large, identical to the arguments used in Section 8.10 for bounding the convergence rate of A_2^* , we know $I \lesssim o(\aleph^4)$ provided that $K^{-\varphi} L_n^\varphi \log(n_*(K \log p)^{-1}) = o(1)$, $\lambda_*^2 (n_*^{-1}K)^{\iota/(2\iota+2)} (\log p)^{-1} = o(1)$, $K \gtrsim L_n$,

$K(\log n_*)^{\chi_1} = o(n_*)$ with $\chi_1 = \max\{(2\iota + 2\gamma + \iota\gamma)/(\iota\gamma), 2 + 2/\iota, 1 + 2/\gamma\}$, and $\log p = o(\min\{(n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}, (n_*K^{-1})^{\chi_2}\})$ with $\chi_2 = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2), \gamma/(\gamma + 4), 1/3\}$. Also, identical to (43),

$$\begin{aligned} \text{II} &\leq p \sum_{i,j=1}^p \{\mathbb{E}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}|^4)\}^{1/2} \{\mathbb{P}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > 4\alpha\aleph)\}^{1/2} \\ &\quad + p \sum_{i,j=1}^p \{\mathbb{E}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}|^4)\}^{1/2} \{\mathbb{P}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > (\beta - \alpha)\aleph)\}^{1/2}. \end{aligned}$$

Notice that $\max_{i,j \in [p]} \mathbb{E}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}|^4) \lesssim n_*^{1/2} \leq p^{2c_*}$ with $c_* = 1/(4\kappa)$, where κ is specified in the beginning of Section 3. Since $\alpha = \beta/2$, repeating the arguments used in Section 8.10 for bounding the convergence rate of A_2^* again, it holds that $\text{II} \lesssim p^{3+c_*} [\max_{i,j \in [p]} \mathbb{P}\{|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > (\beta - \alpha)\aleph\}]^{1/2} \lesssim o(\aleph^2)$ if $K^{-\varphi} L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$, $\lambda_*^2(n_*^{-1}K)^{\iota/(2\iota+2)}(\log p)^{-1} = o(1)$, $K \gtrsim L_n$, $K(\log n_*)^{\chi_1} = o(n_*)$ with $\chi_1 = \max\{(2\iota + 2\gamma + \iota\gamma)/(\iota\gamma), 2 + 2/\iota, 1 + 2/\gamma\}$, $\log p = o(\min\{(n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}, (n_*K^{-1})^{\chi_2}\})$ with $\chi_2 = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2), \gamma/(\gamma + 4), 1/3\}$. Therefore, $\mathbb{E}\{(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|^2)\} \leq \text{I} + \text{II} \lesssim o(\aleph^2)$, which implies $\sup_{\mathcal{P}_4} \mathbb{E}(\|\widehat{\Sigma}_u^{\text{jump,thre}} - \Sigma_u\|_2^2) \lesssim c_p^2(n_*^{-1}K \log p)^{1-q}$. Analogously, in the case with $\varphi = \infty$, we have $\sup_{\mathcal{P}_4} \mathbb{E}(\|\widehat{\Sigma}_u^{\text{jump,thre}} - \Sigma_u\|_2^2) \lesssim c_p^2(n_*^{-1}K \log p)^{1-q}$ provided that $K > L_n$, $\lambda_*^2(n_*^{-1}K)^{\iota/(2\iota+2)}(\log p)^{-1} = o(1)$, $K(\log n_*)^{\chi_1} = o(n_*)$ with $\chi_1 = \max\{(2\iota + 2\gamma + \iota\gamma)/(\iota\gamma), 2 + 2/\iota, 1 + 2/\gamma\}$ and $\log p = o(\min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(2\iota + 2), \gamma/(\gamma + 4), 1/3\})$. We complete the proof of Part (ii) in Theorem 7. \square

8.14. Proof of Theorem 8

For sufficiently large n , $\min_{k \in [n_{i,j}]} N_{i,j,k} = \Delta_K + 1$ and $\max_{k \in [n_{i,j}]} N_{i,j,k} = 2\Delta_K + 2$. Analogously to (48), we have that

$$\begin{aligned} |\varpi_{i,j}| &\leq \frac{1}{2n_{i,j}} \sum_{k=1}^{n_{i,j}} \sum_{K \leq |\ell-k| \leq K+\Delta_K} \left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |U_{i,s}| dM_{i,s} \right) \left(\int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |U_{j,s}| dM_{j,s} \right) \\ &\lesssim \left(K \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} |U_{j,s}| dM_{j,s} \right) \times \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s} \\ &:= Q_{i,j}^J \cdot \frac{1}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s}. \end{aligned} \tag{61}$$

For any constant $d \geq 1$ and $i \in [p]$, define the event $\mathcal{E}_{i,d} = \{\sup_{0 \leq s \leq T} |U_{i,s}| > d\}$. Let $(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} = \int_{t_{i,j,k} \wedge t_{i,j,\ell}}^{t_{i,j,k} \vee t_{i,j,\ell}} dM_{j,s}$ for any $k \in [n_{i,j}]$ and $t_{i,j,\ell} \in S_{i,j,k}$. Recall $(\Delta M_{i,\cdot})_{i,j} = \int_{t_{i,j,1}}^{t_{i,j,n_{i,j}}} dM_{i,s}$. By Assumption 7, it holds that

$$\begin{aligned} \mathbb{P}\left(\frac{Q_{i,j}^J}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s} \geq v\right) &\leq \mathbb{P}\left(\frac{Q_{i,j}^J}{2n_{i,j}} \sum_{\ell=1}^{n_{i,j}-1} \int_{t_{i,j,\ell}}^{t_{i,j,\ell+1}} |U_{i,s}| dM_{i,s} \geq v, \mathcal{E}_{i,d}^c, \mathcal{E}_{j,d}^c\right) + \mathbb{P}(\mathcal{E}_{i,d}) + \mathbb{P}(\mathcal{E}_{j,d}) \\ &\lesssim \mathbb{P}\left\{d^2 K \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} (\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} \cdot (\Delta M_{i,\cdot})_{i,j} \geq 2n_{i,j}v\right\} + \exp(-Cd) \end{aligned} \tag{62}$$

for any $v > 0$. By Bonferroni inequality, we have that

$$\mathbb{P}\left\{\max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} (\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} \geq v\right\} \lesssim n_* \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} \geq v\} \tag{63}$$

for any $v > 0$. Recall that $\xi = \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} |t_{i,j,\ell} - t_{i,j,k}| \asymp n_*^{-1}K$. By Assumption 8(ii), it holds that $\max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)}\} \lesssim \lambda_* n_*^{-1}K$. If $\lambda_* n_*^{-1}K = o(1)$, by Markov's inequality, Assumption 8 implies that

$$\begin{aligned} \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} \geq v\} &= \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} - \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)}\} \geq v - \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)}\}\} \\ &\lesssim \mathbb{P}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} - \mathbb{E}\{(\Delta M_{j,\cdot})_{j,i}^{(k,\ell)}\} \geq v/2\} \lesssim \exp(-Cv) \exp(C\lambda_* n_*^{-1}K) \lesssim \exp(-Cv) \end{aligned}$$

for any $v \gg \lambda_* n_*^{-1}K$. By (63), we have $\mathbb{P}\{\max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} (\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} \geq v\} \lesssim n_* \exp(-Cv)$ for any $v \gg \lambda_* n_*^{-1}K$. Together with (50), for any $\delta \in (0, 1)$, it holds that

$$\begin{aligned} &\mathbb{P}\left\{d^2 K \max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} (\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} \cdot (\Delta M_{i,\cdot})_{i,j} \geq 2n_{i,j}v\right\} \\ &\lesssim \mathbb{P}\left\{\max_{k \in [n_{i,j}]} \max_{t_{i,j,\ell} \in S_{i,j,k}} (\Delta M_{j,\cdot})_{j,i}^{(k,\ell)} \geq \xi^{-\delta}\right\} + \mathbb{P}\{d^2 K (\Delta M_{i,\cdot})_{i,j} \geq 2n_{i,j}\xi^\delta v\} \end{aligned}$$

$$\lesssim n_* \exp(-C\xi^{-\delta}) + \exp(-Cd^{-2}\xi^{\delta-1}v)$$

for any $v \gg \xi^{1-\delta}d^2\lambda_*$. Together with (61) and (62), $\max_{i,j \in [p]} \mathbb{P}(|\varpi_{i,j}| \geq v) \lesssim n_* \exp(-C\xi^{-\delta}) + \exp(-Cd^{-2}\xi^{\delta-1}v) + \exp(-Cd')$ for any $v \gg \xi^{1-\delta}d^2\lambda_*$. In order to make $\max_{i,j \in [p]} |\varpi_{i,j}| = O_p(\xi^{1/2} \log^{1/2} p)$, it suffices to require $\log p = o[\min\{\xi^{-\delta}, d^{-4}\xi^{2\delta-1}, d'\}]$, $\lambda_*^2 \xi^{1-2\delta} d^4 (\log p)^{-1} = o(1)$ and $\log n_* = o(\xi^{-\delta})$. To make p diverge as fast as possible, we select $\delta = \iota/(3\iota + 4)$ and $d' \asymp \xi^{-\delta}$. Then if $K(\log n_*)^{(3\iota+4)/\iota} = o(n_*)$, $\max_{i,j \in [p]} \mathbb{P}(|\varpi_{i,j}| \geq v) \lesssim \exp\{-C(n_*K^{-1})^{(2\iota+2)/(3\iota+4)}v\} + \exp\{-C(n_*K^{-1})^{\iota/(3\iota+4)}\}$ for any $v \gg (n_*^{-1}K)^{(2\iota+2)/(3\iota+4)}\lambda_*$.

Analogous to (58), by Assumptions 7 and 8, it holds that

$$\max_{i,j \in [p]} \mathbb{E}(|\varpi_{i,j}|^m) \lesssim \max_{i,j \in [p]} \max_{k \in [n_{i,j}]} \max_{\ell \in S_{i,j,k}} \mathbb{E} \left\{ \left(\int_{\ell_{i,j,k} \wedge \ell_{i,j,\ell}}^{\ell_{i,j,k} \vee \ell_{i,j,\ell}} |U_{i,s}| dM_{i,s} \right)^{2m} \right\} \lesssim 1$$

for any fixed positive integer m . We complete the proof of Theorem 8. \square

8.15. Proof of Theorem 9

Write $\aleph = (n_*^{-1}K \log p)^{1/2}$. Recall $\hat{\sigma}_{u,i,j}^{\text{jump}} = \hat{\sigma}_{u,i,j} - \varpi_{i,j}$. By Theorem 8 and the proof of Section 8.10, for some sufficiently large constant $\alpha_* > 0$, if $\lambda_*^2(n_*^{-1}K)^{\iota/(3\iota+4)}(\log p)^{-1} = o(1)$, we have

$$\mathbb{P} \left(\max_{i,j \in [p]} |\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| \geq \alpha_* \aleph \right) \lesssim p^2 \max_{i,j \in [p]} \mathbb{P}(|\hat{\sigma}_{u,i,j}^{\text{jump}} - \sigma_{u,i,j}| > \alpha_* \aleph / 2) + p^2 \max_{i,j \in [p]} \mathbb{P}(|\varpi_{i,j}| \geq \alpha_* \aleph / 2) = o(\aleph^4)$$

provided that $\log p = o[\min\{(n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}, (n_*K^{-1})^{\chi^*}\}]$, $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$ and $K \gtrsim L_n$, where $\chi^* = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(3\iota + 4), \gamma/(\gamma + 4)\}$. Hence, Part (i) holds. For any $C > 0$, by Markov's inequality, we have $\mathbb{P}(\|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2 \geq Cc_p\aleph^{1-q}) \leq C^{-2}c_p^{-2}\aleph^{2(q-1)}\mathbb{E}(\|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2)$. In the sequel, we will show $\mathbb{E}(\|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim c_p^2\aleph^{2(1-q)}$. Based on this result, we know Part (ii) holds. For each $i, j \in [p]$, we define the event $A_{i,j} = \{|\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j}| \leq 4 \min\{|\sigma_{u,i,j}|, \alpha \aleph\}\}$ with some constant $\alpha > 0$, and $d_{i,j} = (\hat{\sigma}_{u,i,j}^{\text{thre}} - \sigma_{u,i,j})I(A_{i,j}^c)$. As shown in (40) and (41), $\mathbb{E}(\|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim \mathbb{E}\{(\max_{i \in [p]} \sum_{j=1}^p |d_{i,j}|)^2\} + c_p^2\aleph^{2(1-q)}$ and

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

Recall $\hat{\sigma}_{u,i,j}^{\text{thre}} = \hat{\sigma}_{u,i,j}I(|\hat{\sigma}_{u,i,j}| \geq \beta\aleph)$. By (42), $I \leq p \sum_{i,j=1}^p \sigma_{u,i,j}^2 \mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| \geq (4\alpha - \beta)\aleph\}$. Notice that $\max_{i,j \in [p]} \mathbb{E}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}|^4\} \lesssim n_*^{1/2}$. Identical to (43), $II \lesssim pn_*^{1/4} \sum_{i,j=1}^p \{\mathbb{P}(|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > 4\alpha\aleph)\}^{1/2} + pn_*^{1/4} \sum_{i,j=1}^p [\mathbb{P}\{|\hat{\sigma}_{u,i,j} - \sigma_{u,i,j}| > (\beta - \alpha)\aleph\}]^{1/2}$. Selecting $\alpha = \beta/2$ for some sufficiently large $\beta > 0$, we have $I + II \lesssim o(\aleph^2)$ provided that $\lambda_*^2(n_*^{-1}K)^{\iota/(3\iota+4)}(\log p)^{-1} = o(1)$, $\log p = o[\min\{(n_*L_n^{-2}K)^{\varphi/(3\varphi+2)}, (n_*K^{-1})^{\chi^*}\}]$, $K^{-\varphi}L_n^\varphi \log\{n_*(K \log p)^{-1}\} = o(1)$ and $K \gtrsim L_n$, where $\chi^* = \min\{\iota\gamma/(2\iota + 2\gamma + \iota\gamma), \iota/(3\iota + 4), \gamma/(\gamma + 4)\}$. Thus $\mathbb{E}(\|\hat{\Sigma}_u^{\text{thre}} - \Sigma_u\|_2^2) \lesssim c_p^2\aleph^{2(1-q)}$. We complete the proof of Theorem 9. \square

8.16. Proof of Lemma 1

Write $S_k = \sum_{t=1}^k z_t$. Note that $\max_{t \in [\bar{n}]} \text{Var}(z_t) < \infty$. We will apply the Fuk-Nagaev inequality (Rio, 2017, Theorem 6.2) to bound the tail probability of $\max_{k \in [\bar{n}]} |S_k|$. Define $\alpha^{-1}(u) = \sum_{k=1}^u I\{u < \alpha(k)\}$. Since $\alpha(k) \leq a_1 \exp(-a_2 \tilde{L}_n^{-\varphi} |k - m|_\varphi^+)$, then $\alpha^{-1}(u) \leq m + a_2^{-1/\varphi} \tilde{L}_n \log^{1/\varphi}(a_1 u^{-1})$ for any $u > 0$. Define $Q(u) = \sup_{t \in [\bar{n}]} Q_t(u)$ with $Q_t(u) = \inf\{x > 0 : \mathbb{P}(|z_t| > x) \leq u\}$. Since $\mathbb{P}(|z_t| > x) \leq b_1 \exp(-b_2 x^r)$, then $Q(u) \leq b_2^{-1/r} \log^{1/r}(b_1 u^{-1})$. Define $R(u) = \alpha^{-1}(u)Q(u)$. We have $R(u) \leq c_1 \tilde{L}_n \log^{1/r_*}(c_2 u^{-1}) + c_1 m \log^{1/r}(c_2 u^{-1})$ with $c_1 = b_2^{-1/r} \max(a_2^{-1/\varphi}, 1)$, $c_2 = \max(a_1, b_1)$ and $r_* = r\varphi/(r + \varphi)$. Since $R(u)$ is a right-continuous and non-increasing function, then its inverse function $H(x) = R^{-1}(x) = \inf\{u : R(u) \leq x\} \leq \inf\{u : c_1 \tilde{L}_n \log^{1/r_*}(c_2 u^{-1}) + c_1 m \log^{1/r}(c_2 u^{-1}) \leq x\} \leq \tilde{c}_1 \exp(-\tilde{c}_2 \tilde{L}_n^{-r_*} x^{r_*}) + \tilde{c}_1 \exp(-\tilde{c}_2 m^{-r} x^r)$ for any $x > 0$ with $\tilde{c}_1 = c_2$ and $\tilde{c}_2 = \min\{(2c_1)^{-r_*}, (2c_1)^{-r}\}$. Write $\tilde{u} = \tilde{c}_1 \exp(-\tilde{c}_2 \tilde{L}_n^{-r_*} x^{r_*}) + \tilde{c}_1 \exp(-\tilde{c}_2 m^{-r} x^r)$. Therefore, $\int_0^{H(x)} Q(u) du \leq \int_0^{\tilde{u}} b_2^{-1/r} \log^{1/r}(b_1 u^{-1}) du \lesssim \int_0^{\tilde{u}} y^{1/r} e^{-y} dy \lesssim b_1^{-1} \tilde{u} \log^{1/r}(b_1 \tilde{u}^{-1}) \lesssim \tilde{u}^{1/2} \{\tilde{u}^{r/2} \log(b_1 \tilde{u}^{-1})\}^{1/r}$. As $x \rightarrow +\infty$, we have $\tilde{u} \rightarrow 0^+$, which implies $\tilde{u}^{r/2} \log(b_1 \tilde{u}^{-1}) \rightarrow 0$. Hence there exists a universal constant $\varepsilon > 0$ such that $\tilde{u}^{r/2} \log(b_1 \tilde{u}^{-1}) < \varepsilon$ for any $\tilde{u} \in (0, 2\tilde{c}_1]$. By the definition of \tilde{u} , we have $\int_0^{H(x)} Q(u) du \lesssim \exp(-C\tilde{L}_n^{-r_*} x^{r_*}) + \exp(-Cm^{-r} x^r)$ for any $x > 0$. Recall $S_{\bar{n}}^2 = \sum_{t_1, t_2=1}^{\bar{n}} |\text{Cov}(z_{t_1}, z_{t_2})|$. By the Fuk-Nagaev inequality, $\mathbb{P}(\max_{k \in [\bar{n}]} |S_k| \geq 4\bar{n}\lambda) \leq 4(1 + \bar{n}^2 \rho^{-1} S_{\bar{n}}^{-2} \lambda^2)^{-\rho/2} +$

$4\lambda^{-1} \int_0^{H(\tilde{n}\lambda/\rho)} Q(u) du \lesssim (1 + \tilde{n}^2 \rho^{-1} s_{\tilde{n}}^{-2} \lambda^2)^{-\rho/2} + \lambda^{-1} \{ \exp(-C\tilde{n}^r \lambda^{r*} \rho^{-r*} \tilde{L}_{\tilde{n}}^{-r*}) + \exp(-Cm^{-r} \tilde{n}^r \lambda^r \rho^{-r}) \}$ for any $\lambda > 0$ and $\rho \geq 1$. We complete the proof of Lemma 1. \square

8.17. Proof of Lemma 2

We first consider the case with $\varphi < \infty$. Define $A_{\tilde{n}}(\alpha, u) = \max\{1, \max_{s \in [\tilde{n}]} \sum_{t=s}^{\tilde{n}} \alpha^{1/u}(t - s)\}$. Notice that $\alpha(k) \leq a_1 \exp\{-a_2(\tilde{L}_{\tilde{n}}^{-1}k)^\varphi\}$ for any integer $k \geq 1$. For any $k \geq 2$, it holds that

$$\begin{aligned} A_{\tilde{n}}\{\alpha, 2(k - 1)\} &\leq 1 + \max_{s \in [\tilde{n}]} \sum_{t=s}^{\tilde{n}} \{\alpha(t - s)\}^{1/(2k-2)} \leq 1 + \sum_{m=0}^{\tilde{n}} \{\alpha(m)\}^{1/(2k-2)} \\ &\leq 1 + a_1 + a_1 \sum_{m=1}^{\tilde{n}} \int_{m-1}^m \exp\{-a_2(2k - 2)^{-1} \tilde{L}_{\tilde{n}}^{-\varphi} x^\varphi\} dx \\ &\leq 1 + a_1 + a_1 \int_0^\infty \exp\{-a_2(2k - 2)^{-1} \tilde{L}_{\tilde{n}}^{-\varphi} x^\varphi\} dx \leq C_*(k - 1)^{1/\varphi} \tilde{L}_{\tilde{n}} \end{aligned}$$

for some constant $C_* > 0$ independent of k . Due to $k^k \leq k!e^k$ for any integer $k \geq 1$, we then have $A_{\tilde{n}}^{k-1}\{\alpha, 2(k - 1)\} \leq C_*^{k-1} \{(k - 1)^{k-1}\}^{1/\varphi} \tilde{L}_{\tilde{n}}^{k-1} \leq (C_* e^{2/\varphi} \tilde{L}_{\tilde{n}})(C_* e^{1/\varphi} \tilde{L}_{\tilde{n}})^{k-2} (k!)^{1/\varphi}$. Let $\Gamma_k(x)$ be the k th order cumulant of the random variable x . By Theorem 4.17 of Saulis and Statulevičius (1991) with $\delta = 1$, we have $|\Gamma_k(S_{\tilde{n}})| \leq (k!)^{2+\tilde{r}+1/\varphi} (\tilde{C} \tilde{L}_{\tilde{n}} H_{\tilde{n}}^2)(\tilde{C} \tilde{L}_{\tilde{n}} H_{\tilde{n}})^{k-2} \tilde{n}$ with $\tilde{C} = 2^{7+2\tilde{r}} C_* e^{2/\varphi}$, which implies $|\Gamma_k(S_{\tilde{n}}/\tilde{n})| = \tilde{n}^{-k} |\Gamma_k(S_{\tilde{n}})| \leq (k!)^{2+\tilde{r}+1/\varphi} (\tilde{C} \tilde{L}_{\tilde{n}} H_{\tilde{n}}^2 \tilde{n}^{-1})(\tilde{C} \tilde{L}_{\tilde{n}} H_{\tilde{n}} \tilde{n}^{-1})^{k-2}$. By Lemma 2.4 of Saulis and Statulevičius (1991), it holds that $\mathbb{P}(|S_{\tilde{n}}| \geq \tilde{n}x) \lesssim \exp(-C\tilde{n} \tilde{L}_{\tilde{n}}^{-1} H_{\tilde{n}}^{-2} x^2) + \exp\{-C(\tilde{n} \tilde{L}_{\tilde{n}}^{-1} H_{\tilde{n}}^{-1} x)^{1/(1+\tilde{r})}\}$ for any $x > 0$, where $\tilde{r} = 1 + \tilde{r} + \varphi^{-1}$. Analogously, when $\varphi = \infty$, we can also show $\mathbb{P}(|S_{\tilde{n}}| \geq \tilde{n}x) \lesssim \exp(-C\tilde{n} \tilde{L}_{\tilde{n}}^{-1} H_{\tilde{n}}^{-2} x^2) + \exp\{-C(\tilde{n} \tilde{L}_{\tilde{n}}^{-1} H_{\tilde{n}}^{-1} x)^{1/(1+\tilde{r})}\}$ for any $x > 0$, where $\tilde{r} = 1 + \tilde{r}$. We complete the proof of Lemma 2. \square

Acknowledgments

We thank the editor, the associate editor and three anonymous referees for their constructive comments and suggestions. Chang and Hu were supported in part by the National Natural Science Foundation of China (Grant Numbers 71991472, 72125008 and 11871401). Chang was also supported by the Center of Statistical Research at Southwestern University of Finance and Economics, China. Liu was supported in part by the National Natural Science Foundation of China (Grant Number 72132008) and the Social Science Foundation of Ministry of Education of China (Grant Number 20YJC790074). This research was undertaken with the assistance of resources provided by the High Performance Computing Center of Wuhan University.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jeconom.2022.06.010>.

References

Ait-Sahalia, Y., Fan, J., Li, Y., 2013. The leverage effect puzzle: Disentangling sources of bias at high frequency. *J. Financ. Econ.* 109, 224–249.
 Ait-Sahalia, Y., Fan, J., Xiu, D., 2010. High frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.* 105, 1504–1517.
 Ait-Sahalia, Y., Jacod, J., 2014. *High-Frequency Financial Econometrics*. Princeton University Press.
 Ait-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. *J. Financ. Econ.* 83, 413–452.
 Ait-Sahalia, Y., Mykland, P.A., 2004. Estimators of diffusions with randomly spaced discrete observations: A general theory. *Ann. Statist.* 32, 2186–2222.
 Ait-Sahalia, Y., Xiu, D., 2017. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Amer. Statist. Assoc.* 201, 384–399.
 Ait-Sahalia, Y., Yu, J., 2009. High frequency market microstructure noise estimates and liquidity measures. *Ann. Appl. Stat.* 3, 422–457.
 Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
 Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shepard, N., 2011. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* 162, 149–169.
 Bickel, P.J., Levina, E., 2008a. Regularized estimation of large covariance matrices. *Ann. Statist.* 36, 199–227.
 Bickel, P.J., Levina, E., 2008b. Covariance regularization by thresholding. *Ann. Statist.* 36, 2577–2604.
 Bolger, N., Laurenceau, J.P., 2013. *Intensive Longitudinal Methods: An Introduction To Diary and Experience Sampling Research*. The Guilford Press.
 Boussama, F., Fuchs, F., Stelzer, R., 2011. Stationary and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Process. Appl.* 121, 2331–2360.
 Bradley, R.C., 2005. Basic properties of strong mixing conditions: A survey and some open questions. *Probability Surveys* 2, 107–144.
 Cai, T.T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* 106, 672–684.
 Cai, T.T., Zhang, C.H., Zhou, H.H., 2010. Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* 38, 2118–2144.
 Cai, T.T., Zhou, H.H., 2012a. Minimax estimation of large covariance matrices under ℓ_1 -norm. *Statistic Sinica* 22, 1319–1378.
 Cai, T.T., Zhou, H.H., 2012b. Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* 40, 2389–2420.
 Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. *Econom. Theory* 18, 17–39.

- Chang, J., Delaigle, A., Hall, P., 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., Wu, Y., 2013. Marginal empirical likelihood and sure independence feature screening. *Ann. Statist.* 41, 2132–2148.
- Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics* 159, 116–133.
- Christensen, K., Oomen, R., Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. *J. Financ. Econ.* 114, 576–599.
- Christensen, K., Podolskij, M., Vetter, M., 2013. On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. *J. Multivariate Anal.* 120, 59–84.
- Da, R., Xiu, D., 2021. Disentangling Autocorrelated Intraday Returns. University of Chicago, Unpublished results.
- Davydov, Y.A., 1968. Convergence of distributions generated by stationary stochastic processes. *Theory Probab. Appl.* 13, 691–696.
- Donoho, D.L., Liu, R.C., 1991. Geometrizing rates of convergence. II. *Ann. Statist.* 19, 633–667.
- Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Amer. Statist. Assoc.* 107, 412–428.
- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
- Hafner, C.M., Preminger, A., 2009. On asymptotic theory for multivariate GARCH models. *J. Multivariate Anal.* 100, 2044–2054.
- Hansen, P.R., Lund, A., 2006. Realized variance and market microstructure noise. *J. Bus. Econom. Statist.* 24, 127–161.
- Jacod, J., Li, Y., Mykland, P., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* 119, 2249–2276.
- Jacod, J., Li, Y., Zheng, X., 2017. Statistical properties of microstructure noise. *Econometrica* 85, 1133–1174.
- Lam, C., Fan, J., 2009. Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* 37, 4254–4278.
- Lam, C., Feng, P., Hu, C., 2017. Nonlinear shrinkage estimation of large integrated covariance matrices. *Biometrika* 104, 481–488.
- Le Cam, L., 1973. Convergence of estimates under dimensionality restrictions. *Ann. Statist.* 1, 38–53.
- Li, Z.M., Linton, O., 2022. A ReMeDI for microstructure noise. *Econometrica* 90 (1), 367–389.
- Li, Y., Xie, S., Zheng, X., 2016. Efficient estimation of integrated volatility incorporating trading information. *J. Econometrics* 195, 33–50.
- Liu, C., Tang, C.Y., 2014. A quasi-maximum likelihood approach for integrated covariance matrix estimation with high frequency data. *J. Econometrics* 180, 217–232.
- Petrov, V.V., 1995. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford University Press.
- Rio, E., 2017. *Asymptotic Theory of Weakly Dependent Random Processes*. Springer.
- Rothman, A.J., Levina, E., Zhu, J., 2009. Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* 104, 177–186.
- Saulis, L., Statulevičius, V.A., 1991. *Limit Theorems for Large Deviations*. Springer.
- Tao, M., Wang, Y., Zhou, H., 2013. Optimal sparse volatility matrix estimation for high dimensional Itô processes with measurement errors. *Ann. Statist.* 41, 1816–1864.
- Ubukata, M., Oya, K., 2009. Estimation and testing for dependence in market microstructure noise. *J. Financ. Econ.* 7, 106–151.
- Varneskov, R.T., 2017. Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels. *Econom. Theory* 33, 1457–1501.
- Wainwright, M.J., 2019. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press.
- Wang, J.L., Chiou, J.M., Müller, H.G., 2016. Functional data analysis. *Annu. Rev. Stat. Appl.* 3, 257–295.
- Wong, K.C., Li, Z., Tewari, A., 2020. Lasso guarantees for β -mixing heavy-tailed time series. *Ann. Statist.* 48, 1124–1142.
- Xia, N., Zheng, X., 2018. On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations. *Ann. Statist.* 46, 500–525.
- Zhang, L., 2011. Estimating covariation: Epps effect, microstructure noise. *J. Econometrics* 160, 33–47.
- Zhang, L., Mykland, P.A., Ait-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high frequency data. *J. Amer. Statist. Assoc.* 100, 1394–1411.
- Zhang, X., Wang, J.L., 2016. From sparse to dense functional data and beyond. *Ann. Statist.* 44, 2281–2321.