Working Paper in Economics

202212

Nonparametric Estimation of Large Spot Volatility Matrices for High-Frequency Financial Data

Ruijun Bu
Degui Li
Oliver Linton
Hanchao Wang

https://www.liverpool.ac.uk/management/people/economics/

Nonparametric Estimation of Large Spot Volatility Matrices for High-Frequency Financial Data

Ruijun Bu[*], Degui Li[†], Oliver Linton[‡], Hanchao Wang[§]

This version: March 16, 2022

Abstract

In this paper, we consider estimating spot/instantaneous volatility matrices of high-frequency data collected for a large number of assets. We first combine classic nonparametric kernel-based smoothing with a generalised shrinkage technique in the matrix estimation for noise-free data under a uniform sparsity assumption, a natural extension of the approximate sparsity commonly used in the literature. The uniform consistency property is derived for the proposed spot volatility matrix estimator with convergence rates comparable to the optimal minimax one. For the high-frequency data contaminated by the microstructure noise, we introduce a localised pre-averaging estimation method in the high-dimensional setting which first pre-whitens data via a kernel filter and then uses the estimation tool developed in the noise-free scenario, and further derive the uniform convergence rates for the developed spot volatility matrix estimator. In addition, we also combine the kernel smoothing with the shrinkage technique to estimate the time-varying volatility matrix of the high-dimensional noise vector, and establish the relevant uniform consistency result. Numerical studies are provided to examine performance of the proposed estimation methods in finite samples.

Keywords: Brownian semi-martingale, Kernel smoothing, Microstructure noise, Sparsity, Spot volatility matrix, Uniform consistency.

[*]Management School, University of Liverpool, UK.
[†]Department of Mathematics, University of York, UK.
[‡]Faculty of Economics, University of Cambridge, Cambridge, UK. The corresponding author, obl20@cam.ac.uk.
[§]Zhongtai Securities Institute for Financial Studies, Shandong University, China.

1 Introduction

Modelling high-frequency financial data is one of the most important topics in financial economics and has received increasing attention in recent decades. Continuous-time econometric models such as the Itô semimartingale are often employed in the high-frequency data analysis. One of the main components in these models is the volatility function or matrix (if there are multiple financial assets). In the low-dimensional setting (with a single or a small number of assets), the realised volatility (or covariance matrix) is often used to estimate the integrated volatility over a fixed time period (e.g., Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002, 2004; Andersen *et al.*, 2003). In practice, it is not uncommon that the high-frequency financial data are contaminated by the market microstructure noise, which leads to biased realised volatility if the noise is ignored in the estimation procedure. Hence, various modification techniques such as the two-scale, pre-averaging and realised kernel have been introduced to account for the microstructure noise and produce consistent volatility estimation when there is a single asset or a small number of assets (e.g., Zhang, Mykland and Aït-Sahalia, 2005; Barndorff-Nielsen *et al.*, 2008; Kalnina and Linton, 2008; Jacod *et al.*, 2009; Podolskij and Vetter, 2009; Christensen, Kinnebrock and Podolskij, 2010; Park, Hong and Linton, 2016). Shephard (2005), Andersen, Bollerslev and Diebold (2010) and Aït-Sahalia and Jacod (2014) provide comprehensive reviews for estimating volatility with high-frequency financial data under various settings.

In recent years, financial economists often have to deal with the situation that there are a large amount of high-frequency financial data collected for a large number of assets. A key issue is to estimate the large volatility structure for these assets which has applications in various areas such as the optimal portfolio choice and risk management. Partly motivated by recent developments in large covariance matrix estimation for low-frequency data in the statistical literature, Wang and Zou (2010), Tao, Wang and Zhou (2013) and Kim, Wang and Zou (2016) estimate the large volatility matrix under an approximate sparsity assumption (Bickel and Levina, 2008); Zheng and Li (2011) and Xia and Zheng (2018) study large volatility matrix estimation using the large-dimensional random matrix theory (Bai and Silverstein, 2010); and Lam and Feng (2018) propose a nonparametric eigenvalue-regularised integrated covariance matrix for high-dimensional asset returns. Given that there often exists co-movement between a large number of assets and the co-movement is driven by some risk factors which can be either observable or latent, Fan, Furger and Xiu (2016), Aït-Sahalia and Xiu (2017), Dai, Lu and Xiu (2019) extend the methodologies developed by Fan, Liao and Mincheva (2011, 2013) for large low-frequency data to estimate the large volatility matrix by imposing a continuous-time factor model structure on the high-dimensional and high-frequency financial data, and Aït-Sahalia and Xiu (2019) study the principal component analysis of high-frequency data and derive the asymptotic distribution for the estimates of the realised eigenvalues, eigenvectors and principal components.

The estimation methodologies in the aforementioned literature rely on the realised volatility (or covariance) matrices, measuring the integrated volatility structure over a fixed time interval. In practice, it is often interesting to further explore the actual spot/instantaneous volatility structure and its dynamic change over certain time interval, which is a particularly important measurement for the financial assets when the market is in a volatile period (say, the global financial crisis or COVID-19 outbreak). For a single financial asset, Fan and Wang (2008) and Kristensen (2010) introduce a kernel-based nonparametric method to estimate the spot volatility function and establish its asymptotic properties including the point-wise and global asymptotic distribution theory and uniform consistency. For the noise-contaminated high-frequency data, Zu and Boswijk (2014) combine the two-scale realised volatility with the kernel-weighted technique to estimate the spot volatility, whereas Kanaya and Kristensen (2016) propose a kernel-weighted pre-averaging spot volatility estimation method. Other nonparametric spot volatility estimation methods can be found in Fan, Fan and Lv (2007) and Figueroa-López and Li (2020). Chapter 8 of Aït-Sahalia and Jacod (2014) reviews some recent developments on spot volatility estimation. It seems straightforward to extend this local nonparametric method to estimate the spot volatility matrix for a small number of assets. However, a further extension to the setting with vast financial assets is non-trivial. There is virtually no work on estimating the vast spot volatility matrix except Kong (2018) which considers estimating large spot volatility matrices and their integrated versions under the continuous-time factor model structure for noise-free high-frequency data.

We consider the large spot volatility matrix estimation problem in two scenarios: (i) noise-free high-frequency data, and (ii) noise-contaminated high-frequency data. In scenario (i), we first use the nonparametric kernel-based smoothing method to estimate the volatility and co-volatility functions as in Fan and Wang (2008) and Kristensen (2010), and then apply a generalised shrinkage to off-diagonal estimated entries. With small off-diagonal entries forced to be zeros, the resulting large spot volatility matrix estimate would be non-degenerate. We derive the uniform consistency property for the proposed spot volatility matrix estimator under a uniform sparsity assumption, which is also adopted by Chen, Xu and Wu (2013), Chen and Leng (2016) and Chen, Li and Linton (2019) in the low-frequency data setting. In particular, the derived uniform convergence rates are comparable to the optimal minimax rate in large covariance matrix estimation (e.g., Cai and Zhou, 2012). The number of assets is allowed to be ultra large in the sense that it can grow at an exponential rate of $1/\Delta$ with Δ being the sampling frequency. In scenario (ii) when the high-frequency data are contaminated by the microstructure noise, we extend a localised pre-averaging estimation method from the low-dimensional setting (e.g., Kanaya and Kristensen, 2016) to the high-dimensional one. Specifically, we first pre-average data via a kernel filter and then apply the same estimation method to the kernel fitted high-frequency data (at pseudo-sampling time points) as in the noise-free scenario (i). The microstructure noise vector is assumed to be weakly correlated and heteroskedastic with the time-varying covariance matrix satisfying the uniform sparsity assumption. We further combine the kernel smoothing with generalised shrinkage to estimate the time-varying noise volatility matrix and derive its uniform convergence property. Some simulation studies are provided to examine the finite-sample performance of the proposed estimation methods.

The rest of the paper is organised as follows. In Section 2, we estimate the large spot volatility matrix in the noise-free high-frequency data setting and give the uniform consistency property. In Section 3, we extend the methodology and theory to the noise-contaminated data setting and further estimate the large noise volatility matrix. Section 4 discusses the spot precision matrix estimation and addresses the asynchronicity issue in the estimation. Section 5 reports the simulation studies. Section 6 concludes the paper. All the mathematical proofs are available in Appendices A and B. Throughout the paper, we let $\|\cdot\|_2$ be the Euclidean norm of a vector; and for a $d \times d$ matrix $\mathbf{A} = (A_{ij})_{d\times d}$, we let $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ be the matrix spectral norm and Frobenius norm, $|\mathbf{A}|_1 = \sum_{i=1}^d \sum_{i=1}^d |A_{ij}|$, $\|\mathbf{A}\|_1 = \max_{1\leqslant j\leqslant d} \sum_{i=1}^d |A_{ij}|$, $\|\mathbf{A}\|_{\infty,q} = \max_{1\leqslant i\leqslant d} \sum_{j=1}^d |A_{ij}|^q$ and $\|\mathbf{A}\|_{\max} = \max_{1\leqslant i\leqslant d} \max_{1\leqslant j\leqslant d} |A_{ij}|$.

2 Estimation with noise-free data

Suppose that $\mathbf{X}_t = (X_{1,t}, \cdots, X_{p,t})^\intercal$ is a p-variate Brownian semi-martingale solving the following stochastic differential equation:

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\mathbf{W}_t, \tag{2.1}$$

where $\mathbf{W}_t = (W_{1,t}, \cdots, W_{p,t})^\intercal$ is a p-dimensional standard Brownian motion, $\boldsymbol{\mu}_t = (\mu_{1,t}, \cdots, \mu_{p,t})^\intercal$ is a p-dimensional drift vector, and $\boldsymbol{\sigma}_t = (\sigma_{ij,t})_{p\times p}$ is a $p \times p$ matrix. The spot volatility matrix of $\mathbf{X}_t$ is defined as

$$\boldsymbol{\Sigma}_t = (\Sigma_{ij,t})_{p\times p} = \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^\intercal. \tag{2.2}$$

Our main interest lies in estimating $\boldsymbol{\Sigma}_t$ when the size p is large. As in Chen, Xu and Wu (2013) and Chen and Leng (2016), we assume that the true spot volatility matrix satisfies the following uniform sparsity condition: $\{\boldsymbol{\Sigma}_t :\ 0 \leqslant t \leqslant T\} \in \mathcal{S}(q, \varpi(p), T)$ which is defined by

$$\left\{ \boldsymbol{\Sigma}_t = [\Sigma_{ij,t}]_{p\times p},\ t \in [0,T] \ \Big|\ \sup_{0\leqslant t\leqslant T} \|\boldsymbol{\Sigma}_t\|_{\infty,q} \leqslant \Lambda\varpi(p) \right\}, \tag{2.3}$$

where $0 \leqslant q < 1$, T is a fixed positive number and Λ is a positive random variable satisfying $\mathsf{E}[\Lambda] \leqslant C_\Lambda < \infty$. This is a natural extension of the approximate sparsity assumption (e.g., Bickel and Levina, 2008; Tao, Wang and Zhou, 2013). The asset prices are assumed to be collected over a fixed time interval $[0, T]$ at $0, \Delta, 2\Delta, \cdots, n\Delta$, where Δ is the sampling frequency and $n = \lfloor T/\Delta \rfloor$ with $\lfloor \cdot \rfloor$ denoting the floor function. In this section, we focus on a simple case of equidistant time points in the high-frequency data collection and will discuss the asynchronicity issue in Section 4.2.

For each $1 \leqslant i, j \leqslant p$, we estimate the spot co-volatility $\Sigma_{ij,t}$ by

$$\widehat{\Sigma}_{ij,t} = \sum_{k=1}^{n} K_h(t_k - t)\Delta X_{i,k}\Delta X_{j,k}, \quad 0 < t < T, \tag{2.4}$$

where $t_k = k\Delta$, $K_h(u) = h^{-1}K(u/h)$, $K(\cdot)$ is a kernel function, h is a bandwidth shrinking to zero and $\Delta X_{i,k} = X_{i,t_k} - X_{i,t_{k-1}}$. A naive method is to estimate the spot volatility matrix $\mathbf{\Sigma}_t$ by $\widehat{\mathbf{\Sigma}}_t$, directly using $\widehat{\Sigma}_{ij,t}$ as its entry. However, this estimate often performs poorly in practice when the number of assets is very large (say, $p > n$). To address this issue, a commonly-used technique is to apply a shrinkage function to $\widehat{\Sigma}_{ij,t}$ when $i \neq j$, forcing very small estimated off-diagonal entries to be zeros. Let $s_\rho(\cdot)$ denote a shrinkage function satisfying the following three conditions: (i) $|s_\rho(u)| \leqslant |u|$ for $u \in \mathcal{R}$; (ii) $s_\rho(u) = 0$ if $|u| \leqslant \rho$; and (iii) $|s_\rho(u) - u| \leqslant \rho$, where ρ is a user-specified tuning parameter. With the shrinkage function, we construct the following nonparametric estimator of $\mathbf{\Sigma}_t$:

$$\widehat{\mathbf{\Sigma}}_t = \left(\widehat{\Sigma}^s_{ij,t}\right)_{p\times p} \quad \text{with} \quad \widehat{\Sigma}^s_{ij,t} = s_{\rho_1(t)}(\widehat{\Sigma}_{ij,t})I(i \neq j) + \widehat{\Sigma}_{ii,t}I(i = j), \tag{2.5}$$

where $\rho_1(t)$ is a tuning parameter which is allowed to change over t and $I(\cdot)$ denotes the indicator function. The above estimation method of the spot volatility matrix can be seen as a natural extension of the recent work on the kernel-based large covariance matrix estimation (e.g., Chen, Xu and Wu, 2013; Chen and Leng, 2016; Chen, Li and Linton, 2019) from the low-frequency data setting to the high-frequency one. We next give some technical assumptions which are needed to derive the uniform convergence property of $\widehat{\mathbf{\Sigma}}_t$.

Assumption 1. *(i) $\{\mu_{i,t}\}$ and $\{\sigma_{ij,t}\}$ are adapted locally bounded processes with continuous sample path.*

(ii) With probability one,

$$\min_{1\leqslant i\leqslant p} \inf_{0\leqslant s\leqslant T} \Sigma_{ii,s} > 0, \quad \min_{1\leqslant i\neq j\leqslant p} \inf_{0\leqslant s\leqslant T} \Sigma^*_{ij,s} > 0,$$

where $\Sigma^*_{ij,s} = \Sigma_{ii,s} + \Sigma_{jj,s} + 2\Sigma_{ij,s}$. For almost all path of the spot covariance process $\{\Sigma_{ij,t}\}$, the m-th derivative (with respect to time), denoted by $\Sigma^{(m)}_{ij,t}$, $m \geqslant 0$, exists and satisfies that

$$\sup_{1\leqslant i,j\leqslant p} \left|\Sigma^{(m)}_{ij,t+\epsilon} - \Sigma^{(m)}_{ij,t}\right| \leqslant B(t, \epsilon)|\epsilon|^\gamma + o(|\epsilon|^\gamma), \quad \epsilon \to 0, \tag{2.6}$$

where $0 < \gamma < 1$, and $B(t, \epsilon)$ is a positive random function slowly varying at $\epsilon = 0$ and continuous *with respect to* t.

Assumption 2. *(i) The kernel* $K(\cdot)$ *is a bounded and Lipschitz continuous function with a compact support* $[-1, 1]$. *In addition,* $\int_{-1}^{1} K(u)du = 1$, $\int_{-1}^{1} u^i K(u)du = 0$, $i = 1, \cdots, \kappa$, *and* $\int_{-1}^{1} |u|^\kappa K(u)du < \infty$.

(ii) The bandwidth h *satisfies that* $h \to 0$ *and* $\frac{h}{\Delta \log(p \vee \Delta^{-1})} \to \infty$.

(iii) Let the time-varying tuning parameter $\rho_1(t)$ *in the generalised shrinkage be chosen as*

$$\rho_1(t) = M(t)\zeta_{\Delta,p}, \quad \zeta_{\Delta,p} = h^{m+\gamma} + \left[\frac{\Delta \log(p \vee \Delta^{-1})}{h}\right]^{1/2},$$

where $M(t)$ *is a positive function satisfying that*

$$0 < \underline{C}_M \leqslant \inf_{0 \leqslant t \leqslant T} M(t) \leqslant \sup_{0 \leqslant t \leqslant T} M(t) \leqslant \overline{C}_M < \infty.$$

Remark 1. Assumption 1 imposes some mild restrictions on the drift and volatility processes. By a typical localisation procedure as in Section 4.4.1 of Jacod and Protter (2012), the local boundedness condition in Assumption 1(i) can be strengthened to the bounded condition over the entire time interval, i.e., with probability one,

$$\max_{1 \leqslant i \leqslant p} \sup_{0 \leqslant s \leqslant T} |\mu_{i,s}| \leqslant C_\mu < \infty, \quad \max_{1 \leqslant i \leqslant p} \sup_{0 \leqslant s \leqslant T} \Sigma_{ii,t} \leqslant C_\Sigma < \infty,$$

which are similar to Assumption A2 in Tao, Wang and Zhou (2013) and Assumptions (A.ii) and (A.iii) in Cai *et al* (2020). Assumption 1(ii) gives the smoothness condition on the spot covariance process, crucial to derive the uniform asymptotic order for the kernel estimation bias. A similar condition is also used by Kristensen (2010) and Zu and Boswijk (2014) in the univariate spot volatility estimation. Note that we allow $\Sigma_{ij,t}$ to be either deterministic or generated by standard stochastic volatility models. When the spot covariance is driven by continuous semimartingales, (2.6) holds with $m = 0$ and $\gamma < 1/2$ (e.g., Ch. V, Exercise 1.20 in Revuz and Yor, 1999). Assumption 2(i) contains some commonly-used conditions for the kernel function. For $\kappa > 2$, $K(\cdot)$ becomes the so-called higher-order kernel, which, together with the condition $m > 2$, leads to reduction of bias order in kernel estimation. Assumptions 2(ii)(iii) impose some mild conditions on the bandwidth and time-varying shrinkage parameter. In particular, when p diverges at a polynomial rate of $1/\Delta$, Assumption 2(ii) reduces to the regular bandwidth restriction. Assumption 2(iii) is comparable to that assumed by Chen and Leng (2016) and Chen, Li and Linton (2019). It is worthwhile to point out that the developed methodology and theory still hold when the time-varying tuning parameter in Assumption 2(iii) is allowed to vary over the (i, j) entries in the spot volatility matrix estimation, which is expected to perform well in finite samples. For example, we construct $\rho_{ij}(t) = \rho(t)\left(\widehat{\Sigma}_{ii,t}\widehat{\Sigma}_{jj,t}\right)^{1/2}$ in the simulation study and shrink the (i,j)-entry to zero if the spot correlation $\widehat{\Sigma}_{ij,t}/\left(\widehat{\Sigma}_{ii,t}\widehat{\Sigma}_{jj,t}\right)^{1/2} \leqslant \rho(t)$.

The following theorem gives the uniform convergence property for the proposed spot volatility matrix estimator $\widehat{\mathbf{\Sigma}}_t$ in the matrix spectral norm.

Theorem 1. *Suppose that Assumptions 1 and 2 are satisfied,* $\{\mathbf{\Sigma}_t : 0 \leqslant t \leqslant T\} \in \mathcal{S}(q, \varpi(p), T)$ *and* $\kappa \geqslant m + \gamma$. *Then we have*

$$\sup_{h\leqslant t\leqslant T-h} \left\|\widehat{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t\right\| = O_P\left(\varpi(p)\zeta_{\Delta,p}^{1-q}\right), \tag{2.7}$$

where $\varpi(p)$ *is defined in (2.3) and* $\zeta_{\Delta,p}$ *is defined in Assumption 2(iii).*

Remark 2. The first term of $\zeta_{\Delta,p}$ is $h^{m+\gamma}$, which is the bias rate due to application of the local smoothing technique. It is smaller than the conventional h^2-rate when $m > 2$ due to Assumption 2(i) on the higher-order kernel function. The second term of $\zeta_{\Delta,p}$ is square root of $\frac{\Delta \log(p \vee \Delta^{-1})}{h}$, the uniform asymptotic rate for the kernel estimation variance component. When p diverges at a polynomial rate of n, $\zeta_{\Delta,p}$ reduces to the uniform convergence rate derived in Theorem 3 of Kristensen (2010) for univariate spot volatility function estimation (see also Kanaya and Kristensen, 2016). The uniform convergence rate in (2.7) is also similar to those obtained by Chen and Leng (2016) and Chen, Li and Linton (2019) in the low-frequency data setting. Note that the dimension p affects the uniform convergence rate via $\varpi(p)$ and $\log(p \vee \Delta^{-1})$ and the estimation consistency may be achieved in the ultra-high dimensional setting when p diverges at an exponential rate of $n = T/\Delta$. Treating (nh) as the "effective" sample size in the local estimation procedure and disregarding the bias order $h^{m+\gamma}$, the uniform convergence rate in (2.7) is comparable to the optimal minimax rate in large covariance matrix estimation (e.g., Cai and Zhou, 2012).

Due to the kernel boundary effect, Theorem 1 only considers the uniform consistency property for the spot volatility matrix estimate $\widehat{\mathbf{\Sigma}}_t$ over the trimmed time interval $[h, T-h]$. In practice, it is often important to investigate the spot volatility structure near the boundary points. For example, when we consider one trading day as a time interval, it is particularly interesting to estimate the spot volatility matrix near the opening and closing times which are peak times in stock market trading. Assume that the underlying spot volatility is driven by continuous semimartingales (e.g., Remark 1), Assumption 1(ii) is satisfied with $m = 0$ and $\gamma < 1/2$. As recommended by Li and Racine (2007), we may replace $K_h(t_k - t)$ in (2.4) by a boundary kernel weight which is defined by

$$K_h^*(t_k - t) = \begin{cases} K_h(t_k - t)/\int_{-t/h}^{1} K(u)du, & 0 \leqslant t < h, \\ K_h(t_k - t), & h \leqslant t \leqslant T-h, \\ K_h(t_k - t)/\int_{-1}^{(T-t)/h} K(u)du, & T-h < t \leqslant T. \end{cases}$$

With the boundary kernel in the proposed spot volatility matrix estimator, we may show that

$$\sup_{0\leqslant t\leqslant T}\left\|\widehat{\boldsymbol{\Sigma}}_t-\boldsymbol{\Sigma}_t\right\|=O_P\left(\varpi(p)\zeta_{\Delta,p,*}^{1-q}\right),$$

where $\zeta_{\Delta,p,*}=h^{\gamma}+\left[\frac{\Delta\log(p\vee\Delta^{-1})}{h}\right]^{1/2}$ with $\gamma<1/2$.

3 Estimation with contaminated high-frequency data

In practice, it is not uncommon that high-frequency financial data are contaminated by the market microstructure noise. The kernel estimation method proposed in Section 2 would be biased if the noise is ignored in the estimation procedure. In this section, we consider the following additive noise structure:

$$\mathbf{Z}_{t_k}=\mathbf{X}_{t_k}+\boldsymbol{\xi}_k=\mathbf{X}_{t_k}+\boldsymbol{\omega}(t_k)\boldsymbol{\xi}_k^*, \tag{3.1}$$

where $t_k=k\Delta$, $k=1,\cdots,n$, $\mathbf{Z}_t=(Z_{1,t},\cdots,Z_{p,t})^{\intercal}$ is a vector of observed asset prices at time t, and $\boldsymbol{\xi}_k=(\xi_{1,k},\cdots,\xi_{p,k})^{\intercal}$ is a p-dimensional vector of noises with nonlinear heteroskedasticity, $\boldsymbol{\omega}(\cdot)=[\omega_{ij}(\cdot)]_{p\times p}$ is a $p\times p$ matrix of deterministic functions, and $\boldsymbol{\xi}_k^*=\left(\xi_{1,k}^*,\cdots,\xi_{p,k}^*\right)^{\intercal}$ independently follows a p-variate identical distribution. The noise structure defined in (3.1) is similar to the setting considered in Kalnina and Linton (2008) which also contains a nonlinear mean function and allows the existence of endogeneity for a single asset. Throughout this section, we assume that $\{\boldsymbol{\xi}_k^*\}$ is independent of the Brownian semimartingale $\{\mathbf{X}_t\}$.

3.1 Estimation of the spot volatility matrix

To account for the microstructure noise and produce consistent volatility matrix estimation, we apply a localised version of the pre-averaging technique as the realised kernel estimate (Barndorff-Nielsen *et al.*, 2008) can be seen as a member of the pre-averaging estimation class whereas the two-scale estimate (Zhang, Mykland and Aït-Sahalia, 2005) can be re-written as the realised kernel estimate with the Bartlett-type kernel (up to the first-order approximation). The pre-averaging method has been studied by Jacod *et al.* (2009), Podolskij and Vetter (2009) and Christensen, Kinnebrock and Podolskij (2010) in estimating the integrated volatility for a single asset and is further extended by Kim, Wang and Zou (2016) and Dai, Lu and Xiu (2019) to the large high-frequency data setting. Kanaya and Kristensen (2016) is among the first to introduce a localised pre-averaging technique to estimate the spot volatility function for a single asset and derive the uniform convergence rate for the developed estimate. A similar technique is also used by Xiao and Linton (2002) to improve convergence of the nonparametric spectral density estimator for time series with general autocorrelation for low-frequency data. The main aim of this section is to extend the localised pre-averaging volatility estimation to the high-dimensional data setting with more flexible noise structure.

We first pre-average the observed high-frequency data via a kernel filter, i.e.,

$$\widetilde{\mathbf{X}}_{\tau} = \frac{T}{n}\sum_{k=1}^{n} L_b(t_k-\tau)\mathbf{Z}_{t_k}, \ \ 0<\tau<T, \tag{3.2}$$

where $L(\cdot)$ is a kernel function and b is a bandwidth. Let $\Delta\widetilde{X}_{i,l} = \widetilde{X}_{i,\tau_l} - \widetilde{X}_{i,\tau_{l-1}}$, where $\widetilde{X}_{i,\tau_l}$ is the i-th component of $\widetilde{\mathbf{X}}_{\tau_l}$ and $\tau_0, \tau_1, \cdots, \tau_N$ are the pseudo-sampling time points in the fixed interval $[0, T]$ with equal distance $\Delta_* = T/N$. Replacing $\Delta X_{i,k}$ by $\Delta\widetilde{X}_{i,l}$ in (2.4), for each $1 \leqslant i, j \leqslant p$, we estimate the spot co-volatility $\Sigma_{ij,t}$ by

$$\widetilde{\Sigma}_{ij,t} = \sum_{l=1}^{N} K_h(\tau_l - t)\Delta\widetilde{X}_{i,l}\Delta\widetilde{X}_{j,l}, \ \ 0<t<T, \tag{3.3}$$

where the kernel weight $K_h(\cdot)$ is defined as in Section 2. Furthermore, to obtain a non-degenerate spot volatility matrix estimate in finite samples when the dimension p is large, as in (2.5), we apply shrinkage to $\widetilde{\Sigma}_{ij,t}$, $1 \leqslant i \neq j \leqslant p$, and subsequently construct

$$\widetilde{\mathbf{\Sigma}}_t = \left(\widetilde{\Sigma}^s_{ij,t}\right)_{p\times p}, \ \ \widetilde{\Sigma}^s_{ij,t} = s_{\rho_2(t)}\left(\widetilde{\Sigma}_{ij,t}\right) I(i\neq j) + \widetilde{\Sigma}_{ii,t} I(i=j), \tag{3.4}$$

where $\rho_2(t)$ is another time-varying shrinkage parameter. We next give some conditions needed to derive the uniform consistency property of $\widetilde{\mathbf{\Sigma}}_t$.

Assumption 3. *(i) Let $\{\xi^*_k\}$ be an independent and identically distributed (i.i.d.) sequence of p-dimensional random vectors. Assume that $\mathsf{E}(\xi^*_{i,k}) = 0$ and*

$$\mathsf{E}\left[\exp\left(s|\mathbf{u}^\intercal \xi^*_k|\right)\right] \leqslant C_\xi < \infty, \ \ 0 < s \leqslant s_0,$$

for any p-dimensional vector $\mathbf{u}$ satisfying $\|\mathbf{u}\|_2 = 1$.

(ii) The deterministic functions $\omega_{ij}(\cdot)$ are bounded uniformly over $i, j \in \{1, \cdots, p\}$, and satisfy that

$$\max_{1\leqslant i\leqslant p}\sup_{0\leqslant t\leqslant T}\sum_{j=1}^{p}\omega^2_{ij}(t) \leqslant C_\omega < \infty.$$

Assumption 4. *(i) The kernel function $L(\cdot)$ is Lipschitz continuous and has a compact support $[-1, 1]$. In addition, $\int_{-1}^{1} L(u)du = 1$.*

(ii) The bandwidth b *and the dimension* p *satisfy that*

$$b \to 0, \quad \frac{\Delta^{2\iota-1} b}{\log(p \vee \Delta^{-1})} \to \infty, \quad p\Delta \exp\{-s\Delta^{-\iota}\} \to 0,$$

where $0 < \iota < 1/2$ *and* $0 < s \leqslant s_0$.

(iii) Let $\nu_{\Delta,p,N} = \sqrt{N \log(p \vee \Delta^{-1})}\left[b^{1/2} + (\Delta^{-1}b)^{-1/2}\right] \to 0$ *and the time-varying tuning parameter* $\rho_2(t)$ *be chosen as* $\rho_2(t) = M(t)\left(\zeta^*_{N,p} + \nu_{\Delta,p,N}\right)$, *where* $M(t)$ *is defined as in Assumption 2(iii) and* $\zeta^*_{N,p}$ *is defined as* $\zeta_{\Delta,p}$ *with* N *replacing* Δ^{-1}.

Remark 3. We allow nonlinear heteroskedasticity on the microstructure noise. The i.i.d. restriction on ξ^*_i may be weakened to some weak dependence conditions (e.g., Kim, Wang and Zou, 2016; Dai, Lu and Xiu, 2019) at the cost of more lengthy proofs. The moment condition in Assumption 3(i) is weaker than the sub-Gaussian condition (e.g., Bickel and Levina, 2008; Tao, Wang and Zhou, 2013) which is commonly used in large covariance matrix estimation when the dimension p is ultra large. The boundedness condition on $\omega_{ij}(\cdot)$ in Assumption 3(ii) is similar to the local boundedness restriction in Assumption 1(i). Assumption 4(ii) imposes some mild restrictions on b and p, which also imply that there is a trade-off between them. When ι is larger, p diverges at a faster exponential rate of $1/\Delta$ but the bandwidth condition becomes more restrictive. If p is divergent at a polynomial rate of $1/\Delta$, we may let ι be sufficiently close to zero, and then the bandwidth condition reduces to the conventional one as in Assumption 2(ii). The condition $\nu_{\Delta,p,N} \to 0$ in Assumption 4(iii) is crucial to show that the error of the kernel filter $\widetilde{\mathbf{X}}_\tau$ tends to zero asymptotically, whereas the form of the time-varying shrinkage parameter $\rho_2(t)$ is relevant to the uniform convergence rate of $\widetilde{\Sigma}_{ij,t}$ (see Proposition A.2).

Theorem 2. *Suppose that Assumptions 1(i)(ii), 2(i), 3 and 4 are satisfied,* $\kappa \geqslant m + \gamma$ *and Assumption 2(ii) holds with* Δ^{-1} *replaced by* N. *When* $\{\mathbf{\Sigma}_t : 0 \leqslant t \leqslant T\} \in \mathcal{S}(q, \varpi(p), T)$, *we have*

$$\sup_{h \leqslant t \leqslant T-h} \left\| \widetilde{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t \right\| = O_P\left(\varpi(p)\left[\zeta^*_{N,p} + \nu_{\Delta,p,N}\right]^{1-q}\right), \tag{3.5}$$

where $\zeta^*_{N,p}$ *and* $\nu_{\Delta,p,N}$ *are defined in Assumption 4(iii).*

Remark 4. The uniform convergence rate in (3.5) relies on $\varpi(p)$, $\zeta^*_{N,p}$ and $\nu_{\Delta,p,N}$. With the high-frequency data collected at pseudo time points with sampling frequency $\Delta_* = T/N$, the rate $\zeta^*_{N,p}$ is comparable to $\zeta_{\Delta,p}$ for the noise-free kernel estimator in Section 2. The rate $\nu_{\Delta,p,N}$ is due to the error of the kernel filter $\widetilde{\mathbf{X}}_\tau$ in the first step of the local pre-averaging estimation procedure. In particular, when $q = 0$, $\varpi(p)$ is bounded, $b = \Delta^{1/4}$ and $h = N^{-\frac{1}{2(m+\gamma)+1}}$ with $N = \Delta^{-\frac{2(m+\gamma)+1}{2[4(m+\gamma)+1]}}$, the uniform convergence rate in (3.5) becomes $\Delta^{\frac{m+\gamma}{2[4(m+\gamma)+1]}}\sqrt{\log(p \vee \Delta^{-1})}$. Furthermore, if $m = 0$ and $\gamma = 1/2$, the rate is simplified to $\Delta^{1/12}\sqrt{\log(p \vee \Delta^{-1})}$, comparable to those derived by Zu and

Boswijk (2014) and Kanaya and Kristensen (2016) in the univariate high-frequency data setting. The boundary kernel defined in Section 2 is applicable to $\widetilde{\Sigma}_{ij,t}$ defined in (3.3) and the uniform consistency result in (3.5) can be extended to cover the entire interval $[0, T]$.

3.2 Estimation of the time-varying noise volatility matrix

In practice, it is often interesting to further explore the volatility structure of microstructure noise. A recent paper by Chang *et al*. (2021) estimates the constant covariance matrix for high-dimensional noise and derives the optimal convergence rates for the developed estimate. In the present paper, we consider the time-varying noise covariance matrix defined by

$$\boldsymbol{\Omega}(t) = \boldsymbol{\omega}(t)\boldsymbol{\omega}^{\intercal}(t) = [\Omega_{ij}(t)]_{p\times p}, \ \ 0 \leqslant t \leqslant T. \tag{3.6}$$

It is sensible to assume that $\{\boldsymbol{\Omega}(t) : \ 0 \leqslant t \leqslant T\}$ satisfies the uniform sparsity condition as in (2.3). For each $1 \leqslant i, j \leqslant p$, we estimate $\Omega_{ij}(t)$ by the kernel smoothing method:

$$\widehat{\Omega}_{ij}(t) = \frac{\Delta}{2}\sum_{k=1}^{n} K_{h_1}(t_k - t)\Delta Z_{i,t_k}\Delta Z_{j,t_k}, \tag{3.7}$$

where h_1 is a bandwidth and $\Delta Z_{i,t_k} = Z_{i,t_k} - Z_{i,t_{k-1}}$. As in (2.5) and (3.4), we again apply shrinkage to $\widetilde{\Omega}_{ij}(t)$, $1 \leqslant i \neq j \leqslant p$, and construct

$$\widehat{\boldsymbol{\Omega}}(t) = \left[\widehat{\Omega}^{s}_{ij}(t)\right]_{p\times p}, \ \ \widehat{\Omega}^{s}_{ij}(t) = s_{\rho_3(t)}\left(\widehat{\Omega}_{ij}(t)\right) I(i \neq j) + \widehat{\Omega}_{ii}(t) I(i = j), \tag{3.8}$$

where $\rho_3(t)$ is a time-varying shrinkage parameter. To derive the uniform consistency property of $\widehat{\boldsymbol{\Omega}}(t)$, we need to impose stronger moment condition on $\boldsymbol{\xi}^{*}_{k}$ and smoothness restriction on $\Omega_{ij}(\cdot)$.

Assumption 5. *(i) For any* p*-dimensional vector* $\mathbf{u}$ *satisfying* $\|\mathbf{u}\|_2 = 1$,

$$\mathsf{E}\left[\exp\left(s(\mathbf{u}^{\intercal}\boldsymbol{\xi}^{*}_{k})^2\right)\right] \leqslant C^{\star}_{\xi} < \infty, \ \ 0 < s \leqslant s_0.$$

(ii) The m*-th derivative of* $\Omega_{ij}(t)$*, denoted by* $\Omega^{(m)}_{ij}(t)$, $m \geqslant 0$, *exists and satisfies that*

$$\sup_{1\leqslant i,j\leqslant p}\left|\Omega^{(m)}_{ij}(t) - \Omega^{(m)}_{ij}(s)\right| \leqslant C_{\Omega}|t - s|^{\gamma},$$

where C_{Ω} *is a positive constant.*

(iii) The bandwidth h_1 *and the dimension* p *satisfy that*

$$h_1 \to 0, \quad \frac{\Delta^{2\iota_\star - 1} h_1}{\log(p \vee \Delta^{-1})} \to \infty, \quad p\Delta^{-1} \exp\{-s\Delta^{-\iota_\star}/C_\omega\} \to 0,$$

where $0 < \iota_\star < 1/2$, $0 < s \leqslant s_0$ *and* C_ω *is defined in Assumption 3(ii).*

Remark 5. Assumption 5(i) strengthens the moment condition in Assumption 3(i) and is equivalent to the sub-Gaussian condition, see Assumption A1 in Tao, Wang and Zhou (2013). The smoothness condition in Assumption 5(ii) is similar to (2.6), crucial to derive the asymptotic order of the kernel estimation bias. The restrictions on h_1 and p in Assumption 5(iii) are similar to those in Assumption 4(ii), allowing the dimension p to be divergent to infinity at an exponential rate of $1/\Delta$.

In the following theorem, we state the uniform consistency result for $\widehat{\mathbf{\Omega}}(t)$ with convergence rate comparable to that in Theorem 1.

Theorem 3. *Suppose that Assumptions 1, 2(i), 3 and 5 are satisfied,* $\kappa \geqslant m + \gamma$ *and Assumption 2(ii) holds when* $\rho_1(t)$, $\zeta_{\Delta,p}$ *and* h *are replaced by* $\rho_3(t)$, $\delta_{\Delta,p}$ *and* h_1, *respectively, where* $\delta_{\Delta,p} = h_1^{m+\gamma} + \left[\frac{\Delta \log(p \vee \Delta^{-1})}{h_1}\right]^{1/2}$. *When* $\{\mathbf{\Omega}(t) : 0 \leqslant t \leqslant T\} \in \mathcal{S}(q, \varpi(p), T)$, *we have*

$$\sup_{h_1 \leqslant t \leqslant T - h_1} \left\| \widehat{\mathbf{\Omega}}(t) - \mathbf{\Omega}(t) \right\| = O_P\left(\varpi(p)\delta_{\Delta,p}^{1-q}\right). \tag{3.9}$$

Remark 6. If the bandwidth parameter h_1 in (3.7) is the same as h in (2.4), we may find that the uniform convergence rate $O_P\left(\varpi(p)\delta_{\Delta,p}^{1-q}\right)$ would be the same as that in Theorem 1. Treating (nh_1) as the "effective" sample size and disregarding the bias order, we may show that the uniform convergence rate in (3.9) is comparable to the optimal minimax rate derived by Chang *et al*. (2021) for the constant noise covariance matrix estimation. Meanwhile, the uniform consistency result in Theorem 3 can be extended to cover the entire interval $[0, T]$ by using the boundary kernel.

4 Discussion and extension

In this section, we discuss estimation of the spot precision matrix and address the asynchronicity issue which is common when multiple asset returns are collected.

4.1 Estimation of the spot precision matrix

The spot precision matrix of high-frequency data defined as inverse of the spot volatility matrix, plays an important role in dynamic optimal portfolio choice. In the low-frequency data setting, estimation of large precision matrices has been extensively studied in the literature and various estimation techniques such as penalised likelihood (Lam and Fan, 2009), graphical Danzig selector (Yuan, 2010) and CLIME (Cai, Liu and Luo, 2011) have been introduced. In the high-frequency data setting, Cai *et al* (2020) estimate the precision matrix defined as inverse of the integrated volatility matrix, derive the relevant asymptotic properties under various scenarios and apply the estimated precision matrix to minimum variance portfolio estimation. In this section, we consider estimating the large spot precision matrix under a uniform sparsity assumption. Specifically, assume that model (3.1) holds and that the large spot precision matrix $\mathbf{\Lambda}_t := \mathbf{\Sigma}_t^{-1}$ satisfies $\{\mathbf{\Lambda}_t : 0 \leqslant t \leqslant T\} \in \mathcal{S}_*(q, \varpi_*(p), T)$ which is defined by

$$\left\{ \mathbf{\Lambda}_t = [\Lambda_{ij,t}]_{p\times p}, t \in [0,T] \;\Big|\; \mathbf{\Lambda}_t \succ 0,\ \sup_{0\leqslant t\leqslant T} \|\mathbf{\Lambda}_t\|_1 \leqslant C_\Lambda,\ \sup_{0\leqslant t\leqslant T} \|\mathbf{\Lambda}_t\|_{\infty,q} \leqslant \varpi_*(p) \right\}, \tag{4.1}$$

where "$\mathbf{\Lambda} \succ 0$" denotes that $\mathbf{\Lambda}$ is positive definite and C_Λ is a positive constant.

We next apply Cai, Liu and Luo (2011)'s constrained ℓ_1 minimisation or CLIME method to estimate the spot precision matrix $\mathbf{\Lambda}_t$. The estimate is defined as

$$\widetilde{\mathbf{\Lambda}}_t = \arg\min_{\mathbf{\Lambda}} |\mathbf{\Lambda}|_1 \quad \text{subject to} \quad \left\| \widetilde{\mathbf{\Sigma}}_t \mathbf{\Lambda} - \mathbf{I}_p \right\|_{\max} \leqslant \rho_4(t),$$

where $\widetilde{\mathbf{\Sigma}}_t = \left(\widetilde{\Sigma}_{ij,t}\right)_{p\times p}$ with $\widetilde{\Sigma}_{ij,t}$ defined in (3.3), $\mathbf{I}_p$ is a $p \times p$ identity matrix, and $\rho_4(t)$ is a time-varying tuning parameter. The final CLIME estimate of $\mathbf{\Lambda}_t$ is obtained by further symmetrising $\widetilde{\mathbf{\Lambda}}_t$. Suppose that Assumptions 1, 2(i), 3 and 4(i)(ii) are satisfied and Assumption 4(iii) holds with $\rho_2(t)$ replaced by $\rho_4(t)$. Using Proposition A.2 in Appendix A and following the proof of Theorem 6 in Cai, Liu and Luo (2011), we may show that

$$\sup_{h\leqslant t\leqslant T-h} \left\| \widetilde{\mathbf{\Lambda}}_t - \mathbf{\Lambda}_t \right\| = O_P\left(\varpi_*(p) \left[\zeta_{N,p}^* + \nu_{\Delta,p,N} \right]^{1-q} \right). \tag{4.2}$$

4.2 The asynchronicity issue

In Sections 2 and 3, we consider a very special sampling scheme: the high-frequency data are synchronised with equally spaced time points between 0 and T. Such a setting simplifies exposition and facilitates proofs of the uniform consistency properties. However, in practice, it is often the case that a large number of assets are traded at times that are not synchronised. This may induce volatility matrix estimation bias and possibly result in the so-called Epps effect (e.g., Epps, 1979). We next deal with the asynchronicity problem and discuss modifications of the estimation techniques and theory developed in the previous sections.

Assume that the i-th asset price is collected at $t_1^i, \cdots, t_{n_i}^i$, which are non-equidistant time points over $[0, T]$. To address this asynchronicity issue, we may adopt a synchronisation scheme before implementing the large spot volatility matrix estimation method proposed in Sections 2 and 3. Commonly-used synchronisation schemes include the generalised sampling time (Aït-Sahalia, Fan and Xiu, 2010), refresh time (Barndorff-Nielsen *et al.*, 2011) and previous tick (Zhang, 2011). We next propose an alternative technique by slightly amending the localised pre-averaging estimation in (3.2) to jointly tackle the asynchronicity and noise contamination issues. Replace the kernel filter in (3.2) by

$$\widetilde{\mathbf{X}}_\tau^* = \left(\widetilde{X}_{1,\tau}^*, \cdots, \widetilde{X}_{p,\tau}^*\right)^\intercal \quad \text{with} \quad \widetilde{X}_{i,\tau}^* = \sum_{k=1}^{n_i} L_b(t_k^i - \tau) Z_{i,t_k^i} \left(t_k^i - t_{k-1}^i\right), \tag{4.3}$$

and then use $\widetilde{\mathbf{X}}_\tau^*$ in the kernel smoothing (3.3). It is easy to verify that the uniform consistency result (3.5) still holds by slightly modifying the proofs of Lemma B.1 and Proposition A.2 in Appendix B and imposing mild restrictions on the time points t_k^i.

The time-varying noise covariance matrix estimation also needs to be modified when large high-frequency data are non-synchronised. As in Chang *et al.* (2021), we let $\mathcal{T}_i = \left\{t_1^i, t_2^i, \cdots, t_{n_i}^i\right\}$ be the set of time points at which we observe the contaminated asset prices, and denote

$$\mathcal{T}_{ij} = \mathcal{T}_i \cap \mathcal{T}_j = \left\{t_1^{ij}, t_2^{ij}, \cdots, t_{n_{ij}}^{ij}\right\},$$

where n_{ij} is the cardinality of $\mathcal{T}_{ij}$. Then, we modify the kernel estimate in (3.7) as follows,

$$\widetilde{\Omega}_{ij}(t) = \frac{1}{2} \sum_{k=1}^{n_{ij}} K_{h_1}\left(t_k^{ij} - t\right) \Delta Z_{i,t_k^{ij}} \Delta Z_{j,t_k^{ij}} \left(t_k^{ij} - t_{k-1}^{ij}\right),$$

where $t_0^{ij} = 0$. In contrast to $\widehat{\Omega}_{ij}(t)$, t_k, Z_{i,t_k} and Δ in (3.7) are now replaced by t_k^{ij}, $Z_{i,t_k^{ij}}$ and $t_k^{ij} - t_{k-1}^{ij}$, respectively. We subsequently apply the shrinkage to $\widetilde{\Omega}_{ij}(t)$ when $i \neq j$ and obtain the final estimate of $\mathbf{\Omega}(t)$. Assuming $\max_{1 \leqslant i,j \leqslant p} \max_{1 \leqslant k \leqslant n_{ij}} \left(t_k^{ij} - t_{k-1}^{ij}\right) \to 0$ and letting $n_\circ = \min_{1 \leqslant i,j \leqslant p} n_{ij}$, we may similarly derive the uniform consistency property as in (3.9) but with Δ replaced by $n_\circ^{-1}$.

5 Monte-Carlo simulation studies

In this section, we report the Monte-Carlo simulation studies to assess the numerical performance of the proposed large spot volatility matrix and time-varying noise volatility matrix estimation methods. Both synchronous and asynchronous high-frequency data are simulated in the studies.

5.1 The simulation setup

We generate the noise-contaminated high-frequency data according to model (3.1), where $\boldsymbol{\omega}(t)$ is taken as the Cholesky decomposition of the noise covariance matrix $\boldsymbol{\Omega}(t) = [\Omega_{ij}(t)]_{p\times p}$, $\boldsymbol{\xi}_k^* = \left(\xi_{1,k}^*, \cdots, \xi_{p,k}^*\right)^\intercal$ is an independent p-dimensional random vector of cross-sectionally independent standard normal random variables, the latent return process $\mathbf{X}_t$ of p assets is generated from the following zero-drift model:

$$d\mathbf{X}_t = \boldsymbol{\sigma}_t d\mathbf{W}_t^X, \;\; t \in [0, T], \tag{5.1}$$

as in Wang and Zou (2010), $\mathbf{W}_t^X = \left(W_{1,t}^X, \cdots, W_{p,t}^X\right)^\intercal$ is a standard p-dimensional Brownian motion, and $\boldsymbol{\sigma}_t$ is chosen as the Cholesky decomposition of the spot covariance matrix $\boldsymbol{\Sigma}_t = (\Sigma_{ij,t})_{p\times p}$. In the simulation, we consider the volatility matrix estimation over the time interval of a full trading day, and set the sampling interval to be 15 seconds, i.e., $\Delta = 1/(252 \times 6.5 \times 60 \times 4)$, to generate synchronous data. We consider three structures in $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Omega}(t)$: "banding", "block-diagonal", and "exponentially decaying". Following Wang and Zou (2010), we generate the diagonal elements of $\boldsymbol{\Sigma}_t$ from the following geometric Ornstein-Uhlenbeck model (see also Barndorff-Nielsen and Shephard, 2002):

$$d\log\Sigma_{ii,t} = -0.6\,(0.157 + \log\Sigma_{ii,t})\,dt + 0.25dW_{i,t}^{\Sigma}, \;\; W_{i,t}^{\Sigma} = \iota_i W_{i,t}^X + \sqrt{1-\iota_i^2}W_{i,t}^*,$$

where $\mathbf{W}_t^* = \left(W_{1,t}^*, \cdots, W_{p,t}^*\right)^\intercal$ is a standard p-dimensional Brownian motion independent of $\mathbf{W}_t^X$, and ι_i is a random number generated uniformly between -0.62 and -0.30, reflecting the leverage effects. The diagonal elements of $\boldsymbol{\Omega}(t)$ are defined as daily cyclical deterministic functions of time,

$$\Omega_{ii}(t) = c_i \left\{ \frac{1}{2} [\cos(2\pi t/T) + 1] \times (\overline{\omega} - \underline{\omega}) + \underline{\omega} \right\},$$

where $\overline{\omega} = 1$ and $\underline{\omega} = 0.1$ reflect the observation by Kalnina and Linton (2008) that the noise level is high at both the opening and the closing times of a trading day and is low in the middle of the day, and the scalar c_i controls the noise ratio for each asset which is chosen to match the highest noise ratio considered by Wang and Zou (2010). As in Barndorff-Nielsen and Shephard (2002, 2004), we define a continuous-time stochastic process κ_t^{Σ} by

$$\begin{aligned} \kappa_t^{\Sigma} &= \frac{e^{2\kappa_t}-1}{e^{2\kappa_t}+1}, \quad d\kappa_t = 0.03\,(0.64-\kappa_t)\,dt + 0.118\kappa_t dW_t^{\kappa}, \\ W_t^{\kappa} &= \sqrt{0.96}W_t^{\diamond} - 0.2\sum_{i=1}^{p} W_{i,t}^{X}/\sqrt{p} \end{aligned}$$

where $W_t^{\diamond}$ is a standard univariate Brownian motion independent of $\mathbf{W}_t^X$ and $\mathbf{W}_t^*$. Let

$$\kappa_t^{\Omega} = \frac{\overline{\kappa}-\underline{\kappa}}{2}\left[\cos\left(2\pi t/T\right)+1\right]+\underline{\kappa},$$

where $\overline{\kappa} = 0.5$ and $\underline{\kappa} = -0.5$. We will use κ_t^{Σ} and κ_t^{Ω} to define the off-diagonal elements in $\mathbf{\Sigma}_t$ and $\mathbf{\Omega}(t)$, respectively, which are specified as follows.

- Banding structure for $\mathbf{\Sigma}_t$ and $\mathbf{\Omega}(t)$: The off-diagonal elements are defined by

$$\Sigma_{ij,t} = \left(\kappa_t^{\Sigma}\right)^{|i-j|}\sqrt{\Sigma_{ii,t}\Sigma_{jj,t}}\cdot I\left(|i-j|\leqslant 2\right),$$

  and

$$\Omega_{ij}(t) = \left(\kappa_t^{\Omega}\right)^{|i-j|}\sqrt{\Omega_{ii}(t)\Omega_{jj}(t)}\cdot I\left(|i-j|\leqslant 2\right),$$

  for $1\leqslant i\neq j\leqslant p$.

- Block-diagonal structure for $\mathbf{\Sigma}_t$ and $\mathbf{\Omega}(t)$: The off-diagonal elements are defined by

$$\Sigma_{ij,t} = \left(\kappa_t^{\Sigma}\right)^{|i-j|}\sqrt{\Sigma_{ii,t}\Sigma_{jj,t}}\cdot I\left((i,j)\in\mathcal{B}\right),$$

$$\Omega_{ij}(t) = \left(\kappa_t^{\Omega}\right)^{|i-j|}\sqrt{\Omega_{ii}(t)\Omega_{jj}(t)}\cdot I\left((i,j)\in\mathcal{B}\right),$$

  for $1\leqslant i\neq j\leqslant p$, where $\mathcal{B}$ is a collection of row and column indices (i,j) located within our randomly generated diagonal blocks [1].

- Exponentially decaying structure for $\mathbf{\Sigma}_t$ and $\mathbf{\Omega}(t)$: The off-diagonal elements are defined by

$$\Sigma_{ij,t} = \left(\kappa_t^{\Sigma}\right)^{|i-j|}\sqrt{\Sigma_{ii,t}\Sigma_{jj,t}}, \quad \Omega_{ij}(t) = \left(\kappa_t^{\Omega}\right)^{|i-j|}\sqrt{\Omega_{ii}(t)\Omega_{jj}(t)}, \quad 1\leqslant i\neq j\leqslant p. \tag{5.2}$$

It is clear that the sparsity condition is not satisfied when the off-diagonal elements of $\mathbf{\Sigma}_t$ and

[1] As in Dai, Lu and Xiu (2019), to generate blocks with random sizes, we fix the largest block size at 20 when $p = 200$ and randomly generate the sizes of the remaining blocks from a random integer uniformly picked between 5 and 20, such that the total size of all blocks is $p = 200$. When $p = 500$, the largest size is 40, and the random integer is uniformly picked between 10 and 40. Block sizes are randomly generated but fixed across all Monte Carlo repetitions.

$\mathbf{\Omega}(t)$ are exponentially decaying as in (5.2). To generate the asynchronous data, we follow Wang and Zou (2010) by randomly deleting 2 observations from every consecutive block of 3 synchronous 15-second observations. Consequently, the average number of asynchronous observations for each asset is equal to one third of the number of synchronous observations. The number of assets p is set as $p = 200$ and 500 and the replication number is $R = 200$

5.2 Volatility matrix estimation

In the simulation studies, we consider the following volatility matrix estimates.

- Noise-free spot volatility matrix estimate $\widehat{\mathbf{\Sigma}}_t$ (for synchronous data). This infeasible estimate serves as a benchmark in comparing the numerical performance of various estimation methods. As in Section 2, we apply the kernel smoothing method to estimate $\Sigma_{ij,t}$ by directly using the latent return process $\mathbf{X}_t$, where the bandwidth is determined by the leave-one-out cross validation. We apply four shrinkage methods to $\widehat{\Sigma}_{ij,t}$ for $i \neq j$: hard thresholding (Hard), soft thresholding (Soft), adaptive LASSO (AL) and smoothly clipped absolute deviation (SCAD). For comparison, we also compute the naive estimate without applying any regularisation technique.

- Noise-contaminated spot volatility matrix estimate $\widetilde{\mathbf{\Sigma}}_t$ (for synchronous data). We combine the kernel smoothing with pre-averaging in Section 3.1 to estimate $\Sigma_{ij,t}$ by using the noise-contaminated process $\mathbf{Z}_t$. As in the noise-free estimation, we apply four shrinkage methods to $\widetilde{\Sigma}_{ij,t}$ for $i \neq j$ and also compute the naive estimate without applying the shrinkage.

- Noise-contaminated spot volatility matrix estimate $\tilde{\mathbf{\Sigma}}_t^*$ (for asynchronous data). This is an extension of $\widetilde{\mathbf{\Sigma}}_t$ defined above to the asynchronous high-frequency data with the modification technique introduced in Section 4.2.

- Time-varying noise volatility matrix estimate $\widehat{\mathbf{\Omega}}(t)$ (for synchronous data). We combine the kernel smoothing with four shrinkage techniques in the estimation as in Section 3.2 and also the naive estimate without shrinkage.

- Time-varying noise volatility matrix estimate $\hat{\mathbf{\Omega}}^*(t)$ (for asynchronous data). This is an extension of $\widehat{\mathbf{\Omega}}(t)$ to the asynchronous high-frequency data with the modification technique introduced in Section 4.2.

The choice of tuning parameter in shrinkage is similar to that in Dai, Lu and Xiu (2019). For example, in the noise-free spot volatility estimate, we set the tuning parameter as $\rho_{ij}(t) = \rho(t)\left(\widehat{\Sigma}_{ii,t}\widehat{\Sigma}_{jj,t}\right)^{1/2}$ where $\rho(t)$ is chosen as the minimum value among the grid of values on $[0, 1]$

such that the shrinkage estimate of the spot volatility matrix is positive definite. To evaluate the estimation performance of $\widehat{\boldsymbol{\Sigma}}_t$, we consider 21 equidistant time points on $[0, T]$ and compute the following Mean Frobenius Loss (MFL) and Mean Spectral Loss (MSL) over 200 repetitions:

$$\begin{aligned} \text{MFL} &= \frac{1}{200}\sum_{m=1}^{200}\left(\frac{1}{21}\sum_{j=1}^{21}\left\|\widehat{\boldsymbol{\Sigma}}_{t_j}^{(m)} - \boldsymbol{\Sigma}_{t_j}^{(m)}\right\|_F\right), \\ \text{MSL} &= \frac{1}{200}\sum_{m=1}^{200}\left(\frac{1}{21}\sum_{j=1}^{21}\left\|\widehat{\boldsymbol{\Sigma}}_{t_j}^{(m)} - \boldsymbol{\Sigma}_{t_j}^{(m)}\right\|\right), \end{aligned}$$

where t_j, $j = 1, 2, \cdots, 21$ are the 21 equal-distant time points on the interval $[0, T]$, and $\widehat{\boldsymbol{\Sigma}}_{t_j}^{(m)}$ and $\boldsymbol{\Sigma}_{t_j}^{(m)}$ are respectively the estimated and true spot volatility matrices at t_j for the m-th repetition. The "MFL" and "MSL" can be similarly defined for $\widetilde{\boldsymbol{\Sigma}}_t$, $\tilde{\boldsymbol{\Sigma}}_t^*$, $\widehat{\boldsymbol{\Omega}}(t)$ and $\hat{\boldsymbol{\Omega}}^*(t)$.

5.3 Simulation results

Table 1 reports the simulation results when the dimension is $p = 200$. The three panels in the table (from top to bottom) report the results where the true volatility matrix structures are banding, block-diagonal, and exponentially decaying, respectively. In each panel, the MFL results are reported on the left, whereas the MSL results are reported on the right. The first three rows of each panel contain the MFL and MSL results for the spot volatility matrix estimation (the first two rows are for synchronous data and the third row is for asynchronous data). The fourth and fifth rows contain the results for the time-varying noise volatility matrix estimation for synchronous and asynchronous data, respectively.

For the noise-free estimate $\widehat{\boldsymbol{\Sigma}}_t$, when the volatility matrix structure is banding, the performance of the four shrinkage estimators are substantially better than that of the naive estimate (without any shrinkage). In particular, the results of the soft thresholding, adaptive LASSO and SCAD are very similar and their MFL and MSL values are approximately one third of those of the naive estimator. Meanwhile, the performance of the hard thresholding is less accurate (despite the much stronger level of shrinking used), but is still much better than the naive estimate. These results suggest that the shrinkage technique is an effective tool in estimating the sparse volatility matrix. Similar results are obtained for the noise-contaminated estimates $\widetilde{\boldsymbol{\Sigma}}_t$ and $\tilde{\boldsymbol{\Sigma}}_t^*$ for both the synchronous and asynchronous data. Unsurprisingly, due to the microstructure noise, the MFL and MSL values of the local pre-averaging estimates are noticeably higher than the corresponding values of the noise-free estimates. The finite-sample convergence is slowed down when the high-frequency data are not synchronised. We next turn the attention to the time-varying noise volatility matrix estimates $\widehat{\boldsymbol{\Omega}}(t)$ and $\hat{\boldsymbol{\Omega}}^*(t)$. As in the spot volatility matrix estimation, the naive method again produces the highest MFL and MSL values. The performance of the four shrinkage estimators are similar with the adaptive LASSO and SCAD being slightly better than the hard and soft thresholding. The simulation results for the block-diagonal and exponentially decaying covariance matrix settings, reported in the middle and bottom panels of Table 1, are fairly close to those for the banding setting. Overall, the results in Table 1 show that the shrinkage methods perform well not only in the sparse covariance matrix settings but also in the non-sparse one (i.e., the exponentially decaying setting).

The simulation results when the dimension is $p = 500$ are reported in Table 2. In general, the results are very similar to those in Table 1, so we omit the detailed discussion and comparison to save the space.

Table 1: Estimation results for the spot volatility and time-varying noise covariance matrices when $p = 200$

				“Banding”								
			Frobenius Norm						Spectral Norm			
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\boldsymbol{\Sigma}}_t$	MFL	14.396	11.407	5.490	4.038	4.830	MSL	3.963	1.799	1.073	0.867	0.987
$\widetilde{\boldsymbol{\Sigma}}_t$	MFL	18.497	12.899	12.196	12.064	12.177	MSL	4.796	2.347	2.260	2.255	2.262
$\widetilde{\boldsymbol{\Sigma}}_t^*$	MFL	21.180	13.234	13.723	13.392	13.768	MSL	6.174	2.375	2.458	2.385	2.474
$\widehat{\boldsymbol{\Omega}}(t)$	MFL	11.714	4.226	4.740	3.237	3.960	MSL	3.281	0.682	1.039	0.571	0.753
$\hat{\boldsymbol{\Omega}}^*(t)$	MFL	38.072	4.640	4.647	4.640	4.646	MSL	6.624	0.663	0.666	0.663	0.665
				“Block-diagonal”								
			Frobenius Norm						Spectral Norm			
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\boldsymbol{\Sigma}}_t$	MFL	14.398	11.277	5.818	4.786	5.424	MSL	4.000	2.293	1.310	1.233	1.386
$\widetilde{\boldsymbol{\Sigma}}_t$	MFL	18.475	12.811	12.192	12.059	12.158	MSL	4.915	2.777	2.663	2.669	2.662
$\widetilde{\boldsymbol{\Sigma}}_t^*$	MFL	21.143	13.141	13.648	13.310	13.693	MSL	6.275	2.805	2.821	2.804	2.827
$\widehat{\boldsymbol{\Omega}}(t)$	MFL	11.713	4.076	4.875	3.240	3.964	MSL	3.274	0.741	1.098	0.606	0.816
$\hat{\boldsymbol{\Omega}}^*(t)$	MFL	38.066	4.520	4.528	4.520	4.526	MSL	6.634	0.736	0.738	0.736	0.737
				“Exponentially decaying”								
			Frobenius Norm						Spectral Norm			
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\boldsymbol{\Sigma}}_t$	MFL	14.402	12.033	6.091	5.287	5.976	MSL	4.078	2.456	1.410	1.348	1.510
$\widetilde{\boldsymbol{\Sigma}}_t$	MFL	18.738	13.464	12.748	12.655	12.739	MSL	4.977	2.934	2.810	2.819	2.815
$\widetilde{\boldsymbol{\Sigma}}_t^*$	MFL	21.454	13.772	14.217	13.914	14.258	MSL	6.313	2.961	2.968	2.958	2.972
$\widehat{\boldsymbol{\Omega}}(t)$	MFL	11.715	4.330	4.860	3.355	4.077	MSL	3.297	0.774	1.085	0.626	0.833
$\hat{\boldsymbol{\Omega}}^*(t)$	MFL	38.098	4.716	4.723	4.717	4.722	MSL	6.672	0.762	0.764	0.762	0.764

The selected bandwidths are $h^* = 90$ for $\widehat{\boldsymbol{\Sigma}}_t$, $h^* = 90$ and $b^* = 4$ for $\widetilde{\boldsymbol{\Sigma}}_t$ and $\widetilde{\boldsymbol{\Sigma}}_t^*$, $h_1^* = 90$ for $\widehat{\boldsymbol{\Omega}}(t)$, and $h_1^* = 250$ for $\hat{\boldsymbol{\Omega}}^*(t)$, where $h^* = h/\Delta$, $b^* = b/\Delta$, and $h_1^* = h_1/\Delta$.

Table 2: Estimation results for the spot volatility and time-varying noise covariance matrices when $p = 500$

		"Banding"										
		Frobenius Norm						Spectral Norm				
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\mathbf{\Sigma}}_t$	MFL	21.971	4.067	5.167	4.916	3.954	MSL	3.907	0.621	0.715	0.698	0.568
$\widetilde{\mathbf{\Sigma}}_t$	MFL	28.479	19.193	18.617	17.930	18.466	MSL	4.767	2.339	2.281	2.228	2.281
$\tilde{\mathbf{\Sigma}}_t^*$	MFL	32.710	20.656	20.445	20.600	20.445	MSL	6.212	2.440	2.427	2.430	2.427
$\widehat{\mathbf{\Omega}}(t)$	MFL	18.269	4.045	4.826	5.532	4.547	MSL	3.307	0.461	0.540	0.675	0.519
$\hat{\mathbf{\Omega}}^*(t)$	MFL	93.263	7.348	7.348	7.348	7.348	MSL	10.724	0.681	0.681	0.681	0.681
		"Block-diagonal"										
		Frobenius Norm						Spectral Norm				
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\mathbf{\Sigma}}_t$	MFL	21.973	5.703	6.429	5.928	5.480	MSL	3.999	0.855	1.134	0.895	0.886
$\widetilde{\mathbf{\Sigma}}_t$	MFL	28.682	19.685	19.155	18.539	19.029	MSL	4.917	2.854	2.782	2.736	2.798
$\tilde{\mathbf{\Sigma}}_t^*$	MFL	32.928	21.080	20.873	21.026	20.873	MSL	6.330	2.962	2.951	2.948	2.950
$\widehat{\mathbf{\Omega}}(t)$	MFL	18.271	4.208	4.935	5.686	4.684	MSL	3.312	0.522	0.603	0.751	0.572
$\hat{\mathbf{\Omega}}^*(t)$	MFL	93.281	7.331	7.331	7.331	7.331	MSL	10.759	0.773	0.773	0.773	0.773
		"Exponentially decaying"										
		Frobenius Norm						Spectral Norm				
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\mathbf{\Sigma}}_t$	MFL	21.973	6.069	6.697	6.120	5.739	MSL	4.035	0.894	1.173	0.927	0.921
$\widetilde{\mathbf{\Sigma}}_t$	MFL	28.867	20.195	19.561	18.950	19.454	MSL	4.938	2.914	2.836	2.788	2.850
$\tilde{\mathbf{\Sigma}}_t^*$	MFL	33.153	21.524	21.371	21.459	21.317	MSL	6.341	3.015	3.003	3.001	3.003
$\widehat{\mathbf{\Omega}}(t)$	MFL	18.275	4.335	5.001	5.763	4.745	MSL	3.322	0.533	0.610	0.757	0.578
$\hat{\mathbf{\Omega}}^*(t)$	MFL	93.287	7.469	7.469	7.469	7.469	MSL	10.783	0.781	0.781	0.781	0.781

The selected bandwidths are $h^* = 240$ for $\widehat{\mathbf{\Sigma}}_t$, $h^* = 240$, $b^* = 4$ for $\widetilde{\mathbf{\Sigma}}_t$, $h^* = 240$, $b^* = 6$ for $\tilde{\mathbf{\Sigma}}_t^*$, $h_1^* = 240$ for $\widehat{\mathbf{\Omega}}(t)$, and $h_1^* = 260$ for $\hat{\mathbf{\Omega}}^*(t)$, where $h^* = h/\Delta$, $b^* = b/\Delta$, and $h_1^* = h_1/\Delta$.

6 Conclusion

In this paper, we have explored the nonparametric estimation methods for large spot volatility matrices under the uniform sparsity assumption. The kernel smoothing combined with the generalised shrinkage technique is proposed to estimate the spot volatility matrix for the noise-free high-frequency data and the uniform convergence rate of the proposed estimate is comparable to the minimax one. This nonparametric estimation method is further combined with the kernel pre-averaging to tackle the noise-contaminated high-frequency data. We also develop the nonparametric estimation methodology and uniform convergence theory for the large time-varying noise volatility matrix. Furthermore, we discuss the spot precision matrix estimation and modify the developed estimation methods to address the asynchronicity issue which is very common when a large number of asset returns are collected. The estimation methodology and theory developed in this paper are applicable to $\mathbf{X}_t$ which is the residual process from regressing returns on observed low-dimensional factors with constant regression coefficients, as the residual estimation error would be dominated by the uniform convergence rates derived in Theorems 1 and 2. The simulation results show that the proposed estimation methods and their modification work well in finite samples for both the synchronous and asynchronous data when the underlying spot volatility matrices are either sparse or non-sparse (with exponentially decaying off-diagonal elements).

Several issues can be further explored. For example, the sparsity assumption imposed on the spot volatility matrix may be too restrictive when assets are highly correlated. There often exist co-movements between these highly-correlated asset returns, which may be modelled by a time-varying factor model (e.g., Kong, 2018; Chen, Mykland and Zhang, 2020). It would be an interesting future topic to extend the nonparametric shrinkage methods developed in this paper to estimate the large spot volatility structure of the high-frequency data satisfying the latent factor structure. It is also worthwhile to further study the spot precision matrix estimation which is briefly discussed in Section 4.1 and explore its application to optimal portfolio choice.

Appendix A: Proofs of the main results

In this appendix, we give the proofs of Theorems 1–3. We start with three propositions on the uniform convergence rates for $\widehat{\Sigma}_{ij,t}$, $\widetilde{\Sigma}_{ij,t}$ and $\widehat{\Omega}_{ij}(t)$. Their proofs are available in Appendix B.

Proposition A.1. *Suppose that Assumptions 1 and 2(i)(ii) are satisfied and let $\kappa \geqslant m+\gamma$, where m and γ are defined in Assumption 1(ii) and κ is defined in Assumption 2(ii). Then, we have*

$$\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\widehat{\Sigma}_{ij,t}-\Sigma_{ij,t}\right| = O_P\left(\zeta_{\Delta,p}\right), \tag{A.1}$$

where $\zeta_{\Delta,p} = h^{m+\gamma}+\left[\frac{\Delta\log(p\vee\Delta^{-1})}{h}\right]^{1/2}$.

Proposition A.2. *Suppose that Assumptions 1, 2(i), 3 and 4(i)(ii) are satisfied, $\kappa \geqslant m+\gamma$ and Assumption 2(ii) holds with Δ^{-1} replaced by N.*

$$\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\widetilde{\Sigma}_{ij,t}-\Sigma_{ij,t}\right| = O_P\left(\zeta^*_{N,p}+\nu_{\Delta,p,N}\right), \tag{A.2}$$

where $\zeta^*_{N,p}$ and $\nu_{\Delta,p,N}$ are defined in Assumption 4(iii).

Proposition A.3. *Suppose that Assumptions 1, 2(i), 3 and 5 are satisfied and* $\kappa \geqslant m+\gamma$*. Then, we have*

$$\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\widehat{\Omega}_{ij}(t)-\Omega_{ij}(t)\right| = O_P\left(\delta_{\Delta,p}\right), \tag{A.3}$$

where $\delta_{\Delta,p} = h_1^{m+\gamma} + \left(\frac{\Delta\log(p\vee\Delta^{-1})}{h_1}\right)^{1/2}$.

Proof of Theorem 1. By the definition of $\widehat{\boldsymbol{\Sigma}}_t^s$ and the property of $s_\rho(\cdot)$, we readily have that

$$\begin{aligned}
& \sup_{h\leqslant t\leqslant T-h}\left\|\widehat{\boldsymbol{\Sigma}}_t^s - \boldsymbol{\Sigma}_t\right\| \\
\leqslant\ & \sup_{h\leqslant t\leqslant T-h}\max_{1\leqslant i\leqslant p}\sum_{j=1}^p\left|\widehat{\Sigma}_{ij,t}^s - \Sigma_{ij,t}\right| \\
=\ & \sup_{h\leqslant t\leqslant T-h}\max_{1\leqslant i\leqslant p}\sum_{j=1}^p\left|s_{\rho_1(t)}\left(\widehat{\Sigma}_{ij,t}\right)I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)-\Sigma_{ij,t}\right| \\
=\ & \sup_{h\leqslant t\leqslant T-h}\max_{1\leqslant i\leqslant p}\sum_{j=1}^p\left|s_{\rho_1(t)}\left(\widehat{\Sigma}_{ij,t}\right)I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)-\Sigma_{ij,t}I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)-\right. \\
& \left.\Sigma_{ij,t}I\left(\left|\widehat{\Sigma}_{ij,t}\right|\leqslant\rho_1(t)\right)\right| \\
\leqslant\ & \sup_{h\leqslant t\leqslant T-h}\max_{1\leqslant i\leqslant p}\sum_{j=1}^p\left|s_{\rho_1(t)}\left(\widehat{\Sigma}_{ij,t}\right)-\widehat{\Sigma}_{ij,t}\right|I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)+ \\
& \sup_{h\leqslant t\leqslant T-h}\max_{1\leqslant i\leqslant p}\sum_{j=1}^p\left|\widehat{\Sigma}_{ij,t}-\Sigma_{ij,t}\right|I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)+ \\
& \sup_{h\leqslant t\leqslant T-h}\max_{1\leqslant i\leqslant p}\sum_{j=1}^p\left|\Sigma_{ij,t}\right|I\left(\left|\widehat{\Sigma}_{ij,t}\right|\leqslant\rho_1(t)\right) \\
=:\ & \Pi_1+\Pi_2+\Pi_3.
\end{aligned} \tag{A.4}$$

Define the event

$$\mathcal{G}(M) = \left\{\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\widehat{\Sigma}_{ij,t}-\Sigma_{ij,t}\right|\leqslant M\zeta_{\Delta,p}\right\}$$

where M is a positive constant. For any small $\epsilon>0$, by (A.1), we may find a sufficiently large constant $M_\epsilon>0$ such that

$$\mathsf{P}\left(\mathcal{G}(M_\epsilon)\right)\geqslant 1-\epsilon \tag{A.5}$$

By property (iii) of the shrinkage function and (A.5), we have

$$\Pi_1 \leqslant \sup_{h\leqslant t\leqslant T-h} \rho_1(t) \left[\max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} I\left(\left|\widehat{\Sigma}_{ij,t}\right| > \rho_1(t)\right)\right]$$

and

$$\Pi_2 \leqslant M_\epsilon \zeta_{\Delta,p} \left[\sup_{h\leqslant t\leqslant T-h} \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} I\left(\left|\widehat{\Sigma}_{ij,t}\right| > \rho_1(t)\right)\right]$$

conditional on the event $\mathcal{G}(M_\epsilon)$. By the reverse triangle inequality and Proposition A.1,

$$\left|\widehat{\Sigma}_{ij,t}\right| \leqslant |\Sigma_{ij,t}| + M_\epsilon \zeta_{\Delta,p}$$

on $\mathcal{G}(M_\epsilon)$. Letting $\underline{C}_M = 2M_\epsilon$ in Assumption 2(iii), as $\{\mathbf{\Sigma}_t : 0 \leqslant t \leqslant T\} \in \mathcal{S}(q, \varpi(p), T)$, we have

$$\begin{aligned}
\Pi_1 + \Pi_2 &\leqslant \zeta_{\Delta,p}(\overline{C}_M + M_\epsilon) \left[\sup_{h\leqslant t\leqslant T-h} \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} I\left(\left|\widehat{\Sigma}_{ij,t}\right| > \underline{C}_M \zeta_{\Delta,p}\right)\right] \\
&\leqslant \zeta_{\Delta,p}(\overline{C}_M + M_\epsilon) \left[\sup_{h\leqslant t\leqslant T-h} \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} I\left(\left|\widehat{\Sigma}_{ij,t}\right| > M_\epsilon \zeta_{\Delta,p}\right)\right] \\
&= O_P\left(\zeta_{\Delta,p}\right) \left[\sup_{h\leqslant t\leqslant T-h} \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} \frac{|\Sigma_{ij,t}|^q}{(M_\epsilon \zeta_{\Delta,p})^q}\right] \\
&= O_P\left(\Lambda\varpi(p)\zeta_{\Delta,p}^{1-q}\right) = O_P\left(\varpi(p)\zeta_{\Delta,p}^{1-q}\right). \qquad (A.6)
\end{aligned}$$

on the event $\mathcal{G}(M_\epsilon)$, where $\overline{C}_M$ is defined in Assumption 2(iii). Note that the events $\left\{\left|\widehat{\Sigma}_{ij,t}\right| \leqslant \rho_1(t)\right\}$ and $\mathcal{G}(M_\epsilon)$ jointly imply that $\left\{|\Sigma_{ij,t}| \leqslant \left(\overline{C}_M + M_\epsilon\right)\zeta_{\Delta,p}\right\}$. Then, we may show that

$$\begin{aligned}
\Pi_3 &\leqslant \sup_{h\leqslant t\leqslant T-h} \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} |\Sigma_{ij,t}| I\left(|\Sigma_{ij,t}| \leqslant \left(\overline{C}_M + M_\epsilon\right)\zeta_{\Delta,p}\right) \\
&\leqslant \left(\overline{C}_M + M_\epsilon\right)^{1-q} \zeta_{\Delta,p}^{1-q} \sup_{h\leqslant t\leqslant T-h} \max_{1\leqslant i\leqslant p} \sum_{j=1}^{p} |\Sigma_{ij,t}|^q \\
&= O_P\left(\Lambda\varpi(p)\zeta_{\Delta,p}^{1-q}\right) = O_P\left(\varpi(p)\zeta_{\Delta,p}^{1-q}\right). \qquad (A.7)
\end{aligned}$$

By (A.6) and (A.7), and letting $\epsilon \to 0$ in (A.5), we complete the proof of Theorem 1. ■

Proof of Theorem 2. The proof is similar to the proof of Theorem 1 with Proposition A.2 replacing Proposition A.1. Details are omitted to save the space. ■

Proof of Theorem 3. The proof is similar to the proof of Theorem 1 with Proposition A.3 replacing

Proposition A.1. Details are omitted to save the space. ■

Appendix B: Proofs of technical results

We next provide the detailed proofs of the propositions stated in Appendix A. As in Remark 1, the local boundedness condition in Assumption 1(i) can be strengthened to the following uniform boundedness condition:

$$\max_{1\leqslant i\leqslant p}\sup_{0\leqslant s\leqslant T}|\mu_{i,s}|\leqslant C_\mu<\infty,\quad \max_{1\leqslant i\leqslant p}\sup_{0\leqslant s\leqslant T}\Sigma_{ii,t}\leqslant C_\Sigma<\infty, \tag{B.1}$$

with probability one. Throughout this appendix, we let C denote a generic positive constant whose value may change from line to line.

Proof of Proposition A.1. Throughout this proof, we let $\zeta^*_{\Delta,p}=\left[\frac{\Delta\log(p\vee\Delta^{-1})}{h}\right]^{1/2}$. By (2.1), we have

$$\begin{aligned}(\Delta X_{i,k})(\Delta X_{j,k}) &= \left(\int_{t_{k-1}}^{t_k}\mu_{i,s}ds+\sum_{l=1}^p\int_{t_{k-1}}^{t_k}\sigma_{il,s}dW_{l,s}\right)\left(\int_{t_{k-1}}^{t_k}\mu_{j,u}du+\sum_{l=1}^p\int_{t_{k-1}}^{t_k}\sigma_{jl,u}dW_{l,u}\right)\\
&= \left(\int_{t_{k-1}}^{t_k}\mu_{i,s}ds\int_{t_{k-1}}^{t_k}\mu_{j,u}du\right)+\left(\int_{t_{k-1}}^{t_k}\sum_{l=1}^p\sigma_{il,s}dW_{l,s}\int_{t_{k-1}}^{t_k}\mu_{j,u}du\right)+\\
&\quad \left(\int_{t_{k-1}}^{t_k}\mu_{i,s}ds\int_{t_{k-1}}^{t_k}\sum_{l=1}^p\sigma_{jl,u}dW_{l,u}\right)+\left(\int_{t_{k-1}}^{t_k}\sum_{l=1}^p\sigma_{il,s}dW_{l,s}\int_{t_{k-1}}^{t_k}\sum_{l=1}^p\sigma_{jl,u}dW_{l,u}\right)\\
&= M_{ij,k}(1)+M_{ij,k}(2)+M_{ij,k}(3)+M_{ij,k}(4).\end{aligned}$$

This leads to the following decomposition for $\widehat{\Sigma}_{ij,t}$:

$$\widehat{\Sigma}_{ij,t}=\sum_{k=1}^n K_h(t_k-t)M_{ij,k}(1)+\sum_{k=1}^n K_h(t_k-t)M_{ij,k}(2)+\sum_{k=1}^n K_h(t_k-t)M_{ij,k}(3)+\sum_{k=1}^n K_h(t_k-t)M_{ij,k}(4).$$

By (B.1) and Assumptions 2(i)(ii), we readily have that

$$\begin{aligned}\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\sum_{k=1}^n K_h(t_k-t)M_{ij,k}(1)\right| &\leqslant \max_{1\leqslant i,j\leqslant p}\max_{1\leqslant k\leqslant n}|M_{ij,k}(1)|\sup_{h\leqslant t\leqslant T-h}\sum_{k=1}^n K_h(t_k-t)\\
&\leqslant C\Delta\sup_{h\leqslant t\leqslant T-h}\Delta\sum_{k=1}^n K_h(t_k-t)\\
&= O_P(\Delta)=o_P\left(\zeta^*_{\Delta,p}\right),\end{aligned} \tag{B.2}$$

as $\Delta \sum_{k=1}^{n} K_h(t_k - t)$ is bounded uniformly over $h \leqslant t \leqslant T - h$.

We next show that

$$\max_{1\leqslant i,j\leqslant p} \sup_{h\leqslant t\leqslant T-h} \left| \sum_{k=1}^{n} K_h(t_k - t) M_{ij,k}(4) - \sum_{k=1}^{n} K_h(t_k - t) \int_{t_{k-1}}^{t_k} \Sigma_{ij,s} ds \right| = O_P\left(\zeta_{\Delta,p}^*\right). \tag{B.3}$$

Let $dX_{i,t}^* = \sum_{l=1}^{p} \sigma_{il,t} dW_{l,t}$, $\Delta X_{i,k}^* = \int_{t_{k-1}}^{t_k} \sum_{l=1}^{p} \sigma_{il,s} dW_{l,s}$ and $X_{i,t}^*$ be adapted to the underlying filtration $(\mathcal{F}_t)_{t\geqslant 0}$. Note that

$$\begin{aligned} M_{ij,k}(4) &= \Delta X_{i,k}^* \Delta X_{j,k}^* = \frac{1}{2}\left[\left(\Delta X_{i,k}^* + \Delta X_{j,k}^*\right)\left(\Delta X_{i,k}^* + \Delta X_{j,k}^*\right) - \left(\Delta X_{i,k}^*\right)^2 - \left(\Delta X_{j,k}^*\right)^2\right] \\ &=: \frac{1}{2}\left[M_{ij,k}^*(4) - \left(\Delta X_{i,k}^*\right)^2 - \left(\Delta X_{j,k}^*\right)^2\right]. \end{aligned}$$

Hence, to show (B.3), it is sufficient to prove that

$$\max_{1\leqslant i\leqslant p} \sup_{h\leqslant t\leqslant T-h} \left| \sum_{k=1}^{n} K_h(t_k - t) \left(\Delta X_{i,k}^*\right)^2 - \sum_{k=1}^{n} K_h(t_k - t) \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds \right| = O_P\left(\zeta_{\Delta,p}^*\right) \tag{B.4}$$

and

$$\max_{1\leqslant i,j\leqslant p} \sup_{h\leqslant t\leqslant T-h} \left| \sum_{k=1}^{n} K_h(t_k - t) M_{ij,k}^*(4) - \sum_{k=1}^{n} K_h(t_k - t) \int_{t_{k-1}}^{t_k} \Sigma_{ij,s}^* ds \right| = O_P\left(\zeta_{\Delta,p}^*\right), \tag{B.5}$$

where $\Sigma_{ij,s}^*$ is defined in Assumption 1(ii).

We next only prove (B.4) as the proof of (B.5) is analogous. Consider covering the interval $[h, T-h]$ by some disjoint intervals $\mathcal{T}_v$ with centre τ_v^* and length $d = h^2 \zeta_{\Delta,p}^*$, $v = 1, 2, \cdots, V$. Observe that

$$\begin{aligned} &\max_{1\leqslant i\leqslant p} \sup_{h\leqslant t\leqslant T-h} \left| \sum_{k=1}^{n} K_h(t_k - t) \left(\Delta X_{i,k}^*\right)^2 - \sum_{k=1}^{n} K_h(t_k - t) \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds \right| \\ \leqslant\ &\max_{1\leqslant i\leqslant p} \max_{1\leqslant v\leqslant V} \left| \sum_{k=1}^{n} K_h(t_k - \tau_v^*) \left(\Delta X_{i,k}^*\right)^2 - \sum_{k=1}^{n} K_h(t_k - \tau_v^*) \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds \right| + \\ &\max_{1\leqslant i\leqslant p} \max_{1\leqslant v\leqslant V} \sup_{t\in\mathcal{T}_v} \left| \sum_{k=1}^{n} \left[K_h(t_k - t) - K_h(t_k - \tau_v^*)\right] \left(\Delta X_{i,k}^*\right)^2 \right| + \\ &\max_{1\leqslant i\leqslant p} \max_{1\leqslant v\leqslant V} \sup_{t\in\mathcal{T}_v} \left| \sum_{k=1}^{n} \left[K_h(t_k - t) - K_h(t_k - \tau_v^*)\right] \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds \right|. \end{aligned} \tag{B.6}$$

As the kernel function has the compact support $[-1,1]$, we have, for any $t \in [h, T-h]$,

$$\sum_{k=1}^{n} K_h(t_k - t)\left[\left(\Delta X_{i,k}^*\right)^2 - \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds\right] = \sum_{k=\lfloor (t-h)/\Delta \rfloor}^{\lfloor (t+h)/\Delta \rfloor} K_h(t_k - t)\left[\left(\Delta X_{i,k}^*\right)^2 - \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds\right].$$

Letting $\mathcal{N}$ be a standard normal random variable, by Lemma 1 in Fan, Li and Yu (2012), we have

$$\mathsf{E}\left(\exp\{\psi(\mathcal{N}^2 - 1)\}\right) \leqslant \exp\left\{2\psi^2\right\} \quad \text{for } |\psi| \leqslant 1/4. \tag{B.7}$$

Following the argument in the proof of Lemma 3 in Fan, Li and Yu (2012) and using (B.7), for $k = \lfloor (\tau_v^* - h)/\Delta \rfloor, \cdots, \lfloor (\tau_v^* + h)/\Delta \rfloor$

$$\begin{aligned} & \mathsf{E}\left(\exp\left\{\theta\left(\Delta^{-1}h\right)^{1/2} K_h(t_k - \tau_v^*)\left[\left(\Delta X_{i,k}^*\right)^2 - \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds\right]\right\} |\mathcal{F}_{t_{k-1}}\right) \\ \leqslant\ & \exp\left\{\frac{2\Delta}{h}\theta^2 C_\Sigma^2 K^2\left(\frac{t_k - \tau_v^*}{h}\right)\right\}, \end{aligned}$$

where θ satisfies that $\left|\theta C_\Sigma (\Delta h^{-1})^{1/2} K\left(\frac{t_k - \tau_v^*}{h}\right)\right| \leqslant 1/4$ and C_Σ is defined in (B.1). Consequently, we have

$$\begin{aligned} & \mathsf{E}\left(\exp\left\{\theta\left(\Delta^{-1}h\right)^{1/2} \sum_{k=1}^{n} K_h(t_k - \tau_v^*)\left[\left(\Delta X_{i,k}^*\right)^2 - \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds\right]\right\}\right) \\ =\ & \mathsf{E}\left(\exp\left\{\theta\left(\Delta^{-1}h\right)^{1/2} \sum_{k=\lfloor (t-h)/\Delta \rfloor}^{\lfloor (t+h)/\Delta \rfloor} K_h(t_k - t)\left[\left(\Delta X_{i,k}^*\right)^2 - \int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds\right]\right\}\right) \\ \leqslant\ & \exp\left\{2\theta^2 C_\Sigma^2 \nu_0\right\}, \end{aligned} \tag{B.8}$$

where $\nu_0 = \int_{-1}^{1} K^2(u) du$. By (B.8), using the Markov inequality and choosing $\theta = \frac{\sqrt{\log(p \vee \Delta^{-1})}}{C_\Sigma^2 \nu_0}$, we can prove that

$$\mathsf{P}\left(\left|\sum_{k=1}^{n} K_h(t_k - \tau_v^*)\left(\Delta X_{i,k}^*\right)^2 - \sum_{k=1}^{n} K_h(t_k - \tau_v^*)\int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds\right| > M\zeta_{\Delta,p}^*\right) \leqslant 2\exp\{-C(M)\log(p \vee \Delta^{-1})\},$$

where $C(M)$ is positive and becomes sufficiently large if we choose M to be large enough. Then, by the Bonferroni inequality, we have

$$\mathsf{P}\left(\max_{1 \leqslant i \leqslant p} \max_{1 \leqslant v \leqslant V}\left|\sum_{k=1}^{n} K_h(t_k - \tau_v^*)\left(\Delta X_{i,k}^*\right)^2 - \sum_{k=1}^{n} K_h(t_k - \tau_v^*)\int_{t_{k-1}}^{t_k} \Sigma_{ii,s} ds\right| > M\zeta_{\Delta,p}^*\right)$$

$$\leqslant \sum_{i=1}^{p}\sum_{v=1}^{V} 2\exp\{-C(M)(\log(p\vee\Delta^{-1}))\} \to 0,$$

where the convergence is due to the fact $pV = o\left(\exp\{C_M \log(p\vee\Delta^{-1})\}\right)$ as V is divergent at a polynomial rate of $1/\Delta$ and $C(M)$ is sufficiently large, which implies that

$$\max_{1\leqslant i\leqslant p}\max_{1\leqslant v\leqslant V}\left|\sum_{k=1}^{n} K_h(t_k-\tau_v^*)\left(\Delta X_{i,k}^*\right)^2 - \sum_{k=1}^{n} K_h(t_k-\tau_v^*)\int_{t_{k-1}}^{t_k}\Sigma_{ii,s}ds\right| = O_P(\zeta_{\Delta,p}^*). \tag{B.9}$$

By the smoothness condition on the kernel function in Assumption 2(i), we have

$$\begin{aligned}
&\max_{1\leqslant i\leqslant p}\max_{1\leqslant v\leqslant V}\sup_{t\in\mathcal{T}_v}\left|\sum_{k=1}^{n}\left[K_h(t_k-t)-K_h(t_k-\tau_v^*)\right]\left(\Delta X_{i,k}^*\right)^2\right| \\
\leqslant\ & \max_{1\leqslant v\leqslant V}\sup_{t\in\mathcal{T}_v}\left|K_h(t_k-t)-K_h(t_k-\tau_v^*)\right| \max_{1\leqslant i\leqslant p}\sum_{k=1}^{n}\left(\Delta X_{i,k}^*\right)^2 \\
=\ & O\left(dh^{-2}\right)\max_{1\leqslant i\leqslant p}\sum_{k=1}^{n}\left(\Delta X_{i,k}^*\right)^2.
\end{aligned}$$

Similar to the proof of (B.9), we may show that

$$\max_{1\leqslant i\leqslant p}\sum_{k=1}^{n}\left(\Delta X_{i,k}^*\right)^2 \leqslant \max_{1\leqslant i\leqslant p}\int_0^T \Sigma_{ii,s}ds + o_P(1) = O_P(1)$$

as T is fixed and $\Sigma_{ii,t}$ is uniformly bounded by C_Σ. Hence, by the choice of d, we have

$$\max_{1\leqslant i\leqslant p}\max_{1\leqslant v\leqslant V}\sup_{t\in\mathcal{T}_v}\left|\sum_{k=1}^{n}\left[K_h(t_k-t)-K_h(t_k-\tau_v^*)\right]\left(\Delta X_{i,k}^*\right)^2\right| = O_P(\zeta_{\Delta,p}^*). \tag{B.10}$$

Analogously, we also have

$$\max_{1\leqslant i\leqslant p}\max_{1\leqslant v\leqslant V}\sup_{t\in\mathcal{T}_v}\left|\sum_{k=1}^{n}\left[K_h(t_k-t)-K_h(t_k-\tau_v^*)\right]\int_{t_{k-1}}^{t_k}\Sigma_{ii,s}ds\right| = O_P(\zeta_{\Delta,p}^*). \tag{B.11}$$

By (B.6) and (B.9)–(B.11), we complete the proof of (B.4).

By (B.2), (B.3) and the Hölder inequality, we have

$$\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\sum_{k=1}^{n}K_h(t_k-t)M_{ij,k}(2)\right|^2$$

$$
\begin{aligned}
&\leqslant \max_{1\leqslant i\leqslant p}\sup_{h\leqslant t\leqslant T-h}\sum_{k=1}^{n}K_h(t_k-t)\left(\Delta X_{i,k}^{*}\right)^2 \max_{1\leqslant j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\sum_{k=1}^{n}K_h(t_k-t)\left(\int_{t_{k-1}}^{t_k}\mu_{j,u}du\right)^2\\
&= O_P(\Delta)\cdot O_P(1)=O_P(\Delta),
\end{aligned}
$$

indicating that

$$
\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\sum_{k=1}^{n}K_h(t_k-t)M_{ij,k}(2)\right| = O_P\left(\Delta^{1/2}\right)=o_P\left(\zeta_{\Delta,p}^{*}\right), \tag{B.12}
$$

and similarly,

$$
\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\sum_{k=1}^{n}K_h(t_k-t)M_{ij,k}(3)\right| = O_P\left(\Delta^{1/2}\right)=o_P\left(\zeta_{\Delta,p}^{*}\right). \tag{B.13}
$$

With (B.2), (B.3), (B.12) and (B.13), we prove that

$$
\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\widehat{\Sigma}_{ij,t}-\sum_{k=1}^{n}K_h(t_k-t)\int_{t_{k-1}}^{t_k}\Sigma_{ij,s}ds\right| = O_P\left(\zeta_{\Delta,p}^{*}\right). \tag{B.14}
$$

On the other hand, by Assumption 1(ii), we may use the m-th order Taylor expansion for $\Sigma_{ij,s}$. Then, using Assumption 2(i) and Lemma 7 in Kristensen (2010), we may show that

$$
\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\sum_{k=1}^{n}K_h(t_k-t)\int_{t_{k-1}}^{t_k}\Sigma_{ij,s}ds-\Sigma_{ij,t}\right| = O_P\left(h^{m+\gamma}\right). \tag{B.15}
$$

Then we prove (A.1) by virtue of (B.14) and (B.15). ■

We next turn to the proof of Proposition A.2, in which a crucial step is to derive a uniform consistency for $\widetilde{X}_{i,\tau}$. The latter is stated in Lemma B.1 below.

Lemma B.1. *Suppose that Assumptions 1(i), 3 and 4(i)(ii) are satisfied. Then we have*

$$
\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}\left|\widetilde{X}_{i,\tau_l}-X_{i,\tau_l}\right| = O_P\left(\sqrt{\log(p\vee\Delta^{-1})}\left[b^{1/2}+\left(\Delta^{-1}b\right)^{-1/2}\right]\right). \tag{B.16}
$$

Proof of Lemma B.1. By the definition of $\widetilde{\mathbf{X}}_\tau$ in (3.2), we write

$$
\widetilde{X}_{i,\tau_l}-X_{i,\tau_l}=\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)Z_{i,t_k}-X_{i,\tau_l}=\Pi_{i,l}(1)+\Pi_{i,l}(2)+\Pi_{i,l}(3)+\Pi_{i,l}(4), \tag{B.17}
$$

where

$$\Pi_{i,l}(1) = \frac{T}{n}\sum_{k=1}^{n} L_b(t_k-\tau_l)\xi_{i,k},$$
$$\Pi_{i,l}(2) = \sum_{k=1}^{n} L_b(t_k-\tau_l)\int_{(k-1)\Delta}^{k\Delta}(X_{i,t_k}-X_{i,s})ds,$$
$$\Pi_{i,l}(3) = \sum_{k=1}^{n}\int_{(k-1)\Delta}^{k\Delta}\left[L_b(t_k-\tau_l)-L_b(s-\tau_l)\right]X_{i,s}ds,$$
$$\Pi_{i,l}(4) = \int_0^T L_b(s-\tau_l)X_{i,s}ds - X_{i,\tau_l}.$$

Let $\nu^*_{\Delta,p}=\left[\frac{\Delta\log(p\vee\Delta^{-1})}{b}\right]^{1/2}$, $\boldsymbol{\omega}_i(t_k)=[\omega_{i1}(t_k),\cdots,\omega_{ip}(t_k)]^\intercal$, and $\boldsymbol{\omega}_{i,*}(t_k)=\boldsymbol{\omega}_i(t_k)/\|\boldsymbol{\omega}_i(t_k)\|$. We first consider $\Pi_{i,l}(1)$. Define

$$\xi^\star_{i,k}=\boldsymbol{\omega}_i^\intercal(t_k)\boldsymbol{\xi}^*_k I\left(|\boldsymbol{\omega}_{i,*}^\intercal(t_k)\boldsymbol{\xi}^*_k|\leqslant\Delta^{-\iota}\right),\quad \xi^\diamond_{i,k}=\boldsymbol{\omega}_i^\intercal(t_k)\boldsymbol{\xi}^*_k I\left(|\boldsymbol{\omega}_{i,*}^\intercal(t_k)\boldsymbol{\xi}^*_k|>\Delta^{-\iota}\right), \tag{B.18}$$

where ι is defined in Assumption 4(ii). Note that

$$\begin{aligned}\Pi_{i,l}(1) &= \frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\boldsymbol{\omega}_i^\intercal(t_k)\boldsymbol{\xi}^*_k\\ &= \frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\left[\xi^\star_{i,k}-\mathsf{E}(\xi^\star_{i,k})\right]+\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\left[\xi^\diamond_{i,k}-\mathsf{E}(\xi^\diamond_{i,k})\right]\end{aligned}$$

as $\mathsf{E}(\xi^\star_{i,k})+\mathsf{E}(\xi^\diamond_{i,k})=0$. By the noise moment condition in Assumption 3(i) and the uniform boundedness condition on $\|\boldsymbol{\omega}_i(t_k)\|$ in Assumption 3(ii), we have

$$\mathsf{E}\left(\left|\xi^\diamond_{i,k}\right|\right)\leqslant C_\omega\cdot\mathsf{E}\left[\left|\boldsymbol{\omega}_{i,*}^\intercal(t_k)\boldsymbol{\xi}^*_k\right|I\left(|\boldsymbol{\omega}_{i,*}^\intercal(t_k)\boldsymbol{\xi}^*_k|>\Delta^{-\iota}\right)\right]=O\left(\Delta^{\iota M^\diamond_\xi}\right)=o\left(\nu^*_{\Delta,p}\right),$$

where $M^\diamond_\xi>0$ is arbitrarily large. Then, by Assumptions 3(i), 4(ii) and the Bonferroni and Markov inequalities, we have, for any $\epsilon>0$,

$$\begin{aligned}&\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}\left|\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\left[\xi^\diamond_{i,k}-\mathsf{E}(\xi^\diamond_{i,k})\right]\right|>\epsilon\nu^*_{\Delta,p}\right)\\ \leqslant\ &\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}\left|\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\xi^\diamond_{i,k}\right|>\frac{1}{2}\epsilon\nu^*_{\Delta,p}\right)\\ \leqslant\ &\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{1\leqslant k\leqslant n}\left|\xi^\diamond_{i,k}\right|>0\right)\end{aligned}$$

$$
\begin{aligned}
&\leqslant \mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{1\leqslant k\leqslant n}\left|\boldsymbol{\omega}_{i,*}^{\intercal}(t_k)\boldsymbol{\xi}_k^*\right| > \Delta^{-\iota}\right)\\
&\leqslant \sum_{i=1}^{p}\sum_{k=1}^{n}\mathsf{P}\left(\left|\boldsymbol{\omega}_{i,*}^{\intercal}(t_k)\boldsymbol{\xi}_k^*\right| > \Delta^{-\iota}\right)\\
&\leqslant pn\exp\{-s\Delta^{-\iota}\}C_{\xi} = o(1)
\end{aligned}
$$

for $0 < s < s_0$, where C_{ξ} is defined in Assumption 3(i). Hence, we have

$$
\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}\left|\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\omega_i(t_k)\left[\xi_{i,k}^{\diamond}-\mathsf{E}(\xi_{i,k}^{\diamond})\right]\right| = o_P\left(\nu_{\Delta,p}^*\right). \tag{B.19}
$$

On the other hand, by Assumptions 3 and 4(i)(ii) as well as the Bernstein inequality for the independent sequence (e.g., Proposition 2.14 in Wainwright, 2019), we may show that

$$
\begin{aligned}
&\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}\left|\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\left[\xi_{i,k}^{\star}-\mathsf{E}(\xi_{i,k}^{\star})\right]\right| > M\nu_{\Delta,p}^*\right)\\
\leqslant\ & \sum_{i=1}^{p}\sum_{l=1}^{N}\mathsf{P}\left(\left|\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\left[\xi_{i,k}^{\star}-\mathsf{E}(\xi_{i,k}^{\star})\right]\right| > M\nu_{\Delta,p}^*\right)\\
=\ & O\left(pN\exp\left\{-C_{\star}(M)\log(p\vee\Delta^{-1})\right\}\right) = o(1),
\end{aligned}
$$

where N diverges to infinity at a polynomial rate of n, $C_{\star}(M)$ is positive and could be sufficiently large by letting M be large enough. Therefore, we have

$$
\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}\left|\frac{T}{n}\sum_{k=1}^{n}L_b(t_k-\tau_l)\left[\xi_{i,k}^{\star}-\mathsf{E}(\xi_{i,k}^{\star})\right]\right| = O_P\left(\nu_{\Delta,p}^*\right). \tag{B.20}
$$

By (B.19) and (B.20), we readily have that

$$
\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}|\Pi_{i,l}(1)| = O_P\left(\nu_{\Delta,p}^*\right). \tag{B.21}
$$

For $\Pi_{i,l}(2)$, we write it as

$$
\begin{aligned}
\Pi_{i,l}(2) &= \sum_{k=1}^{n}L_b(t_k-\tau_l)\int_{(k-1)\Delta}^{k\Delta}\left(\int_s^{k\Delta}\mu_{i,u}du\right)ds + \sum_{k=1}^{n}L_b(t_k-\tau_l)\int_{(k-1)\Delta}^{k\Delta}\left(\int_s^{k\Delta}\sum_{j=1}^{p}\sigma_{ij,u}dW_{j,u}\right)ds\\
&= \Pi_{i,l}(2,1)+\Pi_{i,l}(2,2).
\end{aligned}
$$

By (B.1) and Assumption 4(i), we have

$$\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}|\Pi_{i,l}(2,1)| = O_P(\Delta) = o_P\left(\nu^*_{\Delta,p}\right). \tag{B.22}$$

By the Bonferroni inequality, we may show that, for any $\epsilon > 0$

$$\begin{aligned}
& \mathsf{P}\left(\max_{1\leqslant i\leqslant p}\sup_{(k-1)\Delta\leqslant s\leqslant k\Delta}\left|\int_s^{k\Delta}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right| > \epsilon\nu^*_{\Delta,p}\right)\\
\leqslant\ & \sum_{i=1}^p \mathsf{P}\left(\sup_{(k-1)\Delta\leqslant s\leqslant k\Delta}\left|\int_s^{k\Delta}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right| > \epsilon\nu^*_{\Delta,p}\right)\\
\leqslant\ & \sum_{i=1}^p \mathsf{P}\left(\sup_{(k-1)\Delta\leqslant s\leqslant k\Delta}\left|\int_{(k-1)\Delta}^{s}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right| > \frac{1}{2}\epsilon\nu^*_{\Delta,p}\right).
\end{aligned}\tag{B.23}$$

By the conditional Jensen inequality, we may verify that both $\left\{\left|\int_{(k-1)\Delta}^s\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right|\right\}_{s\geqslant(k-1)\Delta}$ and $\left\{\exp\left(\psi\left|\int_{(k-1)\Delta}^s\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right|\right)\right\}_{s\geqslant(k-1)\Delta}$ are sub-martingales, where $\psi > 0$. Using the moment generating function for the folded normal random variable and (B.1), we have

$$\mathsf{E}\left[\exp\left(\psi\left|\int_{(k-1)\Delta}^{k\Delta}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right|\right)\right] \leqslant \exp\left(\frac{\psi^2\Delta C_\Sigma}{2}\right),$$

where C_Σ is defined in (B.1). Combining the above arguments and using Doob's inequality for sub-martingales, we may show that

$$\begin{aligned}
& \mathsf{P}\left(\sup_{(k-1)\Delta\leqslant s\leqslant k\Delta}\left|\int_{(k-1)\Delta}^{s}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right| > \frac{1}{2}\epsilon\nu^*_{\Delta,p}\right)\\
=\ & \mathsf{P}\left(\sup_{(k-1)\Delta\leqslant s\leqslant k\Delta}\exp\left\{\psi\left|\int_{(k-1)\Delta}^{s}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right|\right\} > \exp\left\{\frac{1}{2}\psi\epsilon\nu^*_{\Delta,p}\right\}\right)\\
\leqslant\ & \exp\left(-\frac{\psi\epsilon\nu^*_{\Delta,p}}{2}\right)\mathsf{E}\left[\exp\left(\psi\left|\int_{(k-1)\Delta}^{k\Delta}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right|\right)\right]\\
\leqslant\ & \exp\left(\frac{\psi^2\Delta C_\Sigma}{2}-\frac{\psi\epsilon\nu^*_{\Delta,p}}{2}\right).
\end{aligned}\tag{B.24}$$

Then, choosing $\psi = \epsilon\nu^*_{\Delta,p}/(2\Delta C_\Sigma)$, by (B.23) and(B.24), we have

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\sup_{(k-1)\Delta\leqslant s\leqslant k\Delta}\left|\int_s^{k\Delta}\sum_{j=1}^p\sigma_{ij,u}dW_{j,u}\right| > \epsilon\nu^*_{\Delta,p}\right)$$

$$\leqslant \quad p \exp\left\{-\frac{(\epsilon \nu^*_{\Delta,p})^2}{8\Delta C_\Sigma}\right\} = O\left(p \exp\left\{-\frac{\epsilon^2}{8C_\Sigma} \cdot \frac{\log(p \vee \Delta^{-1})}{b}\right\}\right) = o(1)$$

for any $\epsilon > 0$, which indicates that

$$\max_{1\leqslant i\leqslant p} \max_{0\leqslant l\leqslant N} |\Pi_{i,l}(2,2)| = o_P\left(\nu^*_{\Delta,p}\right). \tag{B.25}$$

By (B.22) and (B.25), we readily have that

$$\max_{1\leqslant i\leqslant p} \max_{0\leqslant l\leqslant N} |\Pi_{i,l}(2)| = o_P\left(\nu^*_{\Delta,p}\right). \tag{B.26}$$

For $\Pi_{i,l}(3)$, we note that

$$|\Pi_{i,l}(3)| \leqslant \sup_{0\leqslant u\leqslant T} |X_{i,u}| \cdot \sum_{k=1}^{n} \int_{(k-1)\Delta}^{k\Delta} |L_b(t_k - \tau_l) - L_b(s - \tau_l)|\, ds.$$

By Assumption 4(i), we have

$$\max_{0\leqslant l\leqslant N} \sum_{k=1}^{n} \int_{(k-1)\Delta}^{k\Delta} |L_b(t_k - \tau_l) - L_b(s - \tau_l)|\, ds = O\left(\Delta b^{-1}\right). \tag{B.27}$$

On the other hand, by (B.1),

$$\sup_{0\leqslant u\leqslant T} |X_{i,u}| = \sup_{0\leqslant u\leqslant T} \int_0^u |\mu_{i,u}| du + \sup_{0\leqslant u\leqslant T} \left|\int_0^u \sum_{j=1}^{p} \sigma_{ij,u} dW_{j,u}\right| = \sup_{0\leqslant u\leqslant T} \left|\int_0^u \sum_{j=1}^{p} \sigma_{ij,u} dW_{j,u}\right| + O_P(1).$$

Following the proof of (B.25), we may show that

$$\sup_{0\leqslant u\leqslant T} \left|\int_0^u \sum_{j=1}^{p} \sigma_{ij,u} dW_{j,u}\right| = O_P\left(\sqrt{\log(p \vee \Delta^{-1})}\right),$$

indicating that

$$\sup_{0\leqslant u\leqslant T} |X_{i,u}| = O_P\left(\sqrt{\log(p \vee \Delta^{-1})}\right). \tag{B.28}$$

By virtue of (B.27) and (B.28), we prove that

$$\max_{1\leqslant i\leqslant p} \max_{0\leqslant l\leqslant N} |\Pi_{i,l}(3)| = O_P\left(\Delta b^{-1} \sqrt{\log(p \vee \Delta^{-1})}\right) = o_P\left(\nu^*_{\Delta,p}\right). \tag{B.29}$$

Finally, for $\Pi_{i,l}(4)$, we write it as

$$\begin{aligned}\Pi_{i,l}(4) &= \left\{\int_0^T L_b(s-\tau_l)\int_0^s \mu_{i,u}du ds - \int_0^{\tau_l}\mu_{i,u}du\right\} + \\ &\quad \left\{\int_0^T L_b(s-\tau_l)\int_0^s \sum_{j=1}^p \sigma_{ij,u}dW_{j,u}ds - \int_0^{\tau_l}\sum_{j=1}^p \sigma_{ij,u}dW_{j,u}\right\} \\ &=: \Pi_{i,l}(4,1)+\Pi_{i,l}(4,2).\end{aligned}$$

By Assumptions 1(i) and 4(i), we readily have that

$$\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}|\Pi_{i,l}(4,1)| = O_P(b). \tag{B.30}$$

Following the proof of (B.25), we may show that

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{1\leqslant l\leqslant N}\sup_{\tau_l\leqslant s\leqslant \tau_l+b}\left|\int_{\tau_l}^s \sum_{j=1}^p \sigma_{ij,u}dW_{j,u}\right| > M\sqrt{b\log(p\vee\Delta^{-1})}\right)\to 0$$

and

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{1\leqslant l\leqslant N}\sup_{\tau_l-b\leqslant s\leqslant \tau_l}\left|\int_{s}^{\tau_l} \sum_{j=1}^p \sigma_{ij,u}dW_{j,u}\right| > M\sqrt{b\log(p\vee\Delta^{-1})}\right)\to 0$$

when $M>0$ is sufficiently large. Consequently, we have

$$\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}|\Pi_{i,l}(4,2)| = O_P\left(\sqrt{b\log(p\vee\Delta^{-1})}\right). \tag{B.31}$$

Combining (B.30) and (B.31),

$$\max_{1\leqslant i\leqslant p}\max_{0\leqslant l\leqslant N}|\Pi_{i,l}(4)| = O_P\left(\sqrt{b\log(p\vee\Delta^{-1})}\right). \tag{B.32}$$

The proof of (B.16) in Lemma B.1 is completed with (B.21), (B.26), (B.29) and (B.32). ■

Proof of Proposition A.2. By (3.3), we have

$$\begin{aligned}\widetilde{\Sigma}_{ij,t}-\Sigma_{ij,t} &= \sum_{l=1}^N K_h(\tau_l-t)\Delta\widetilde{X}_{i,l}\Delta\widetilde{X}_{j,l}-\Sigma_{ij,t} \\ &= \sum_{l=1}^N K_h(\tau_l-t)\Delta X_{i,l}\Delta X_{j,l}-\Sigma_{ij,t}+\sum_{k=1}^3 \Xi_{ij,t}(k),\end{aligned}$$

where

$$\Xi_{ij,t}(1) = \sum_{l=1}^{N} K_h(\tau_l - t)\Delta X_{i,l}\left(\Delta\widetilde{X}_{j,l} - \Delta X_{j,l}\right),$$
$$\Xi_{ij,t}(2) = \sum_{l=1}^{N} K_h(\tau_l - t)\left(\Delta\widetilde{X}_{i,l} - \Delta X_{i,l}\right)\Delta X_{j,l},$$
$$\Xi_{ij,t}(3) = \sum_{l=1}^{N} K_h(\tau_l - t)\left(\Delta\widetilde{X}_{i,l} - \Delta X_{i,l}\right)\left(\Delta\widetilde{X}_{j,l} - \Delta X_{j,l}\right).$$

By Proposition A.1, we have

$$\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left|\sum_{l=1}^{N} K_h(\tau_l - t)\Delta X_{i,l}\Delta X_{j,l} - \Sigma_{ij,t}\right| = O_P\left(h^{m+\gamma} + \left[\frac{\log(p\vee N)}{Nh}\right]^{1/2}\right). \tag{B.33}$$

By Lemma B.1 and Assumption 2(i), we have

$$\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}|\Xi_{ij,t}(3)| = O_P\left(N\log(p\vee\Delta^{-1})\left[b^{1/2} + \left(\Delta^{-1}b\right)^{1/2}\right]^2\right). \tag{B.34}$$

By Proposition A.1, (B.34) and the Hölder inequality, we have

$$\max_{1\leqslant i,j\leqslant p}\sup_{h\leqslant t\leqslant T-h}\left(|\Xi_{ij,t}(1)| + |\Xi_{ij,t}(2)|\right) = O_P\left(\sqrt{N\log(p\vee\Delta^{-1})}\left[b^{1/2} + \left(\Delta^{-1}b\right)^{1/2}\right]\right). \tag{B.35}$$

The proof of (A.2) in Proposition A.2 is completed by virtue of (B.33)–(B.35). ■

Proof of Proposition A.3. By (3.1) and (3.7), we write

$$\begin{aligned}\widehat{\Omega}_{ij}(t) &= \frac{\Delta}{2}\sum_{k=1}^{n} K_{h_1}(t_k - t)\Delta X_{i,k}\Delta X_{j,k} + \frac{\Delta}{n}\sum_{k=1}^{n} K_{h_1}(t_k - t)\Delta X_{i,k}(\xi_{j,k} - \xi_{j,k-1}) + \\ &\quad \frac{\Delta}{2}\sum_{k=1}^{n} K_{h_1}(t_k - t)(\xi_{i,k} - \xi_{i,k-1})\Delta X_{j,k} + \frac{\Delta}{2}\sum_{k=1}^{n} K_{h_1}(t_k - t)(\xi_{i,k} - \xi_{i,k-1})(\xi_{j,k} - \xi_{j,k-1}) \\ &=: \widehat{\Omega}_{ij,1}(t) + \widehat{\Omega}_{ij,2}(t) + \widehat{\Omega}_{ij,3}(t) + \widehat{\Omega}_{ij,4}(t).\end{aligned}$$

By Proposition A.1, we have

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\widehat{\Omega}_{ij,1}(t)\right| = O_P(\Delta). \tag{B.36}$$

To complete the proof of (A.3), it is sufficient to show

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\widehat{\Omega}_{ij,4}(t)-\Omega_{ij}(t)\right| = O_P\left(\delta_{\Delta,p}\right). \tag{B.37}$$

In fact, combining (B.36) and (B.37), and using the Hölder inequality, we

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left[\left|\widehat{\Omega}_{ij,2}(t)\right|+\left|\widehat{\Omega}_{ij,3}(t)\right|\right] = O_P\left(\Delta^{1/2}\right). \tag{B.38}$$

By virtue of (B.36)–(B.38), we readily have (A.3).

It remains to prove (B.37). We aim to show that

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\xi_{i,k}\xi_{j,k}-\Omega_{ij}(t)\right| = O_P\left(\delta_{\Delta,p}\right), \tag{B.39}$$

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\xi_{i,k-1}\xi_{j,k-1}-\Omega_{ij}(t)\right| = O_P\left(\delta_{\Delta,p}\right), \tag{B.40}$$

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\left(\xi_{i,k}\xi_{j,k-1}+\xi_{i,k-1}\xi_{j,k}\right)\right| = O_P\left(\delta^*_{\Delta,p}\right), \tag{B.41}$$

where $\delta^*_{\Delta,p}=\left[\frac{\Delta\log(p\vee\Delta^{-1})}{h_1}\right]^{1/2}$. To save the space, we only provide the detailed proof of (B.39) as the proofs of (B.40) and (B.41) are similar (with minor modifications).

Note that

$$\begin{aligned}
&\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\xi_{i,k}\xi_{j,k}-\Omega_{ij}(t)\\
=&\left\{\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\left[\xi_{i,k}\xi_{j,k}-\Omega_{ij}(t_k)\right]\right\}+\left\{\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\Omega_{ij}(t_k)-\Omega_{ij}(t)\right\}\\
=:&\ \Upsilon_{ij,1}(t)+\Upsilon_{ij,2}(t).
\end{aligned} \tag{B.42}$$

Let $\chi_{ij,k}=\xi_{i,k}\xi_{j,k}-\Omega_{ij}(t_k)$,

$$\chi^\star_{ij,k}=\chi_{ij,k}I\left(|\chi_{ij,k}|\leqslant\Delta^{-\iota_\star}\right)\quad\text{and}\quad\chi^\diamond_{ij,k}=\chi_{ij,k}-\chi^\star_{ij,k},$$

where $\iota_\star$ is defined in Assumption 5(ii). Observe that

$$\Upsilon_{ij,1}(t)=\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\left[\chi^\star_{ij,k}-\mathsf{E}(\chi^\star_{ij,k})\right]+\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\left[\chi^\diamond_{ij,k}-\mathsf{E}(\chi^\diamond_{ij,k})\right]. \tag{B.43}$$

By Assumptions 3(ii) and 5(i), we have $\mathsf{E}\left[\left|\chi^{\diamond}_{ij,k}\right|\right] = O\left(\Delta^{\iota_\star M_\chi}\right)$ with $M_\chi > 0$ being arbitrarily large. Then, by Assumptions 5(i)(ii) and the Markov inequality, we have that, for any $\epsilon > 0$,

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\left[\chi^{\diamond}_{ij,k}-\mathsf{E}(\chi^{\diamond}_{ij,k})\right]\right| > \epsilon\delta^{*}_{\Delta,p}\right)\\
\leqslant\;& \mathsf{P}\left(\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\chi^{\diamond}_{ij,k}\right| > \frac{1}{2}\epsilon\delta^{*}_{\Delta,p}\right)\\
\leqslant\;& \mathsf{P}\left(\max_{1\leqslant i,j\leqslant p}\max_{1\leqslant k\leqslant n}\left|\chi^{\diamond}_{ij,k}\right| > 0\right) \leqslant \mathsf{P}\left(\max_{1\leqslant i,j\leqslant p}\max_{1\leqslant k\leqslant n}|\chi_{ij,k}| > \Delta^{-\iota_\star}\right)\\
\leqslant\;& \mathsf{P}\left(\max_{1\leqslant i,j\leqslant p}\max_{1\leqslant k\leqslant n}|\xi_{i,k}\xi_{j,k}| > \Delta^{-\iota_\star}-M_\Omega\right) \leqslant \mathsf{P}\left(\max_{1\leqslant i,j\leqslant p}\max_{1\leqslant k\leqslant n}\left(\xi^2_{i,k}+\xi^2_{j,k}\right) > 2(\Delta^{-\iota_\star}-M_\Omega)\right)\\
\leqslant\;& 2\mathsf{P}\left(\max_{1\leqslant i\leqslant p}\max_{1\leqslant k\leqslant n}\xi^2_{i,k} > \Delta^{-\iota_\star}-M_\Omega\right) \leqslant 2\sum_{i=1}^{p}\sum_{k=1}^{n}\mathsf{P}\left(\xi^2_{i,k} > \Delta^{-\iota_\star}-M_\Omega\right)\\
\leqslant\;& 2pn\exp\{-sC_\omega^{-1}\left(\Delta^{-\iota_\star}-M_\Omega\right)\}C^{\star}_{\xi} = o(1)
\end{aligned} \tag{B.44}$$

for $0 < s < s_0$, where $M_\Omega = \max_{1\leqslant i,j\leqslant p}\sup_{0\leqslant t\leqslant T}|\Omega_{ij}(t)| \leqslant C_\omega$, C_ω is defined in Assumption 3(ii) and $C^{\star}_{\xi}$ is defined in Assumption 5(i).

Cover the closed interval $[h_1, T-h_1]$ by some disjoint intervals $\mathcal{T}^{\star}_{l}$, $l = 1,\cdots,V_\star$, with the center $t^{\star}_{l}$ and length $d_\star = h_1^2\delta^{*}_{\Delta,p}\Delta^{\iota_\star}$. By the Lipschitz continuity of $K(\cdot)$ in Assumption 2(i), we have

$$\begin{aligned}
&\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t)\left[\chi^{\star}_{ij,k}-\mathsf{E}(\chi^{\star}_{ij,k})\right]\right|\\
\leqslant\;& \max_{1\leqslant i,j\leqslant p}\max_{1\leqslant l\leqslant V_\star}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t^{\star}_{l})\left[\chi^{\star}_{ij,k}-\mathsf{E}(\chi^{\star}_{ij,k})\right]\right| +\\
&\max_{1\leqslant i,j\leqslant p}\max_{1\leqslant l\leqslant V_\star}\sup_{t\in\mathcal{T}^{\star}_{l}}\left|\Delta\sum_{k=1}^{n}\left[K_{h_1}(t_k-t)-K_{h_1}(t_k-t^{\star}_{l})\right]\left[\chi^{\star}_{ij,k}-\mathsf{E}(\chi^{\star}_{ij,k})\right]\right|\\
\leqslant\;& \max_{1\leqslant i,j\leqslant p}\max_{1\leqslant l\leqslant V_\star}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t^{\star}_{l})\left[\chi^{\star}_{ij,k}-\mathsf{E}(\chi^{\star}_{ij,k})\right]\right| +\\
&O\left(\Delta^{-\iota_\star}\right)\max_{1\leqslant l\leqslant V_\star}\sup_{t\in\mathcal{T}^{\star}_{l}}\Delta\sum_{k=1}^{n}\left|K_{h_1}(t_k-t)-K_{h_1}(t_k-t^{\star}_{l})\right|\\
\leqslant\;& \max_{1\leqslant i,j\leqslant p}\max_{1\leqslant l\leqslant V_\star}\left|\Delta\sum_{k=1}^{n}K_{h_1}(t_k-t^{\star}_{l})\left[\chi^{\star}_{ij,k}-\mathsf{E}(\chi^{\star}_{ij,k})\right]\right| + O_P\left(\delta^{*}_{\Delta,p}\right).
\end{aligned} \tag{B.45}$$

On the other hand, by the Bernstein inequality, we may show that

$$\begin{aligned} & \mathsf{P}\left(\max_{1\leqslant i,j\leqslant p}\max_{1\leqslant l\leqslant V_\star}\left|\Delta\sum_{k=1}^n K_{h_1}(t_k-t_l^\star)\left[\chi_{ij,k}^\star-\mathsf{E}(\chi_{ij,k}^\star)\right]\right|>M\delta_{\Delta,p}^*\right) \\ \leqslant & \sum_{i=1}^p\sum_{j=1}^p\sum_{l=1}^{V_\star}\mathsf{P}\left(\left|\Delta\sum_{k=1}^n K_{h_1}(t_k-t_l^\star)\left[\chi_{ij,k}^\star-\mathsf{E}(\chi_{ij,k}^\star)\right]\right|>M\delta_{\Delta,p}^*\right) \\ = & O\left(p^2V_\star\exp\left\{-C_\diamond(M)\log(p\vee\Delta^{-1})\right\}\right)=o(1), \end{aligned}$$

where $C_\diamond(M)$ is positive and becomes sufficiently large by choosing M to be large enough, and $V_\star$ diverges at a polynomial rate of n. Therefore, we have

$$\max_{1\leqslant i,j\leqslant p}\max_{1\leqslant l\leqslant V_\star}\left|\Delta\sum_{k=1}^n K_{h_1}(t_k-t_l^\star)\left[\chi_{ij,k}^\star-\mathsf{E}(\chi_{ij,k}^\star)\right]\right|=O_P(\delta_{\Delta,p}^*). \tag{B.46}$$

With (B.43)–(B.46), we can prove that

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}|\Upsilon_{ij,1}(t)|=O_P(\delta_{\Delta,p}^*). \tag{B.47}$$

Finally, by the m-th order Taylor expansion of $\Omega_{ij}(\cdot)$ and Assumption 2(i), we have

$$\max_{1\leqslant i,j\leqslant p}\sup_{h_1\leqslant t\leqslant T-h_1}|\Upsilon_{ij,2}(t)|=O\left(h^{m+\gamma}\right). \tag{B.48}$$

By virtue of (B.42), (B.47) and (B.48), we complete the proof of (B.39). ■

References

AÏT-SAHALIA, Y., FAN, J. AND XIU, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* 105, 1504–1517.

AÏT-SAHALIA, Y. AND JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

AÏT-SAHALIA, Y. AND XIU, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics* 201, 384–399.

AÏT-SAHALIA, Y. AND XIU, D. (2019). Principal component analysis of high-frequency data. *Journal of the American Statistical Association* 114, 287–303.

ANDERSEN, T. G. AND BOLLERSLEV, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.

ANDERSEN, T. G., BOLLERSLEV, T. AND DIEBOLD, F. X. (2010). Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics: Tools and Techniques* (Y. Aït-Sahalia and L. P. Hansen, eds.), 67–137.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. AND LABYS, P. (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.

BAI, Z. AND SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics, Springer.

BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B* 64, 253–280.

BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression and correlation in financial economics. *Econometrica* 72, 885–925.

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. AND SHEPHARD, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. AND SHEPHARD, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149–169.

BICKEL, P. AND LEVINA, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics* 36, 2577–2604.

CAI, T. T., HU, J., LI, Y. AND ZHENG, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics* 214, 482–494.

CAI, T. T., LIU, W. AND LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* 106, 594–607.

CAI, T. T. AND ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Annals of Statistics* 40, 2389–2420.

CHANG, J., HU, Q., LIU, C. AND TANG, C. (2021). Optimal covariance matrix estimation for high-dimensional noise in high-frequency data. Working paper available at https://arxiv.org/abs/1812.08217.

CHEN, J., LI, D. AND LINTON, O. (2019). A new semiparametric estimation approach of large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics* 212, 155–176.

CHEN, D., MYKLAND, P. A. AND ZHANG, L. (2020). The five trolls under the bridge: principal component analysis with asynchronous and noisy high frequency data. *Journal of the American Statistical Association* 115, 1960–1977.

CHEN, X., XU, M. AND WU, W. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Annals of Statistics* 41, 2994–3021.

CHEN, Z. AND LENG, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association* 111, 1196–1207.

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

DAI, C., LU, K. AND XIU, D. (2019). Knowing factors or factor loadings, or neither? Evaluating estimators for large covariance matrices with noisy and asynchronous data. *Journal of Econometrics* 208, 43–79.

EPPS, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association* 74, 291–298.

FAN, J., FAN, Y. AND LV, J. (2007). Aggregation of nonparametric estimators for volatility matrix. *Journal of Financial Econometrics* 5, 321–357.

FAN, J., FURGER, A. AND XIU, D. (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. *Journal of Business and Economic Statistics* 34, 489–503.

FAN, J., LI, Y. AND YU, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association* 107, 412–428.

FAN, J., LIAO, Y. AND MINCHEVA, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics* 39, 3320–3356.

FAN, J., LIAO, Y. AND MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* 75, 603–680.

FAN, J. AND WANG, Y. (2008). Spot volatility estimation for high-frequency data. *Statistics and Its Interface* 1, 279–288.

FIGUEROA-LÓPEZ, J. E. AND LI, C. (2020). Optimal kernel estimation of spot volatility of stochastic differential equations. *Stochastic Processes and Their Applications* 130, 4693–4720.

HAYASHI, T. AND YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. AND VETTER, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications* 119, 2249–2276.

JACOD, J. AND PROTTER, P. (2012). *Discretization of Processes*. Springer.

KALNINA, I. AND LINTON, O. (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics* 147, 47–59.

KANAYA, S. AND KRISTENSEN, D. (2016). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory* 32, 861–916.

KIM, D., WANG, Y. AND ZOU, J. (2016). Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Processes and Their Applications* 126, 3527–3577.

KONG, X. (2018). On the systematic and idiosyncratic volatility with large panel high-frequency data. *The Annals of Statistics* 46, 1077–1108.

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

LAM, C. AND FAN, J. (2009). Sparsity and rates of convergence in large covariance matrix estimation. *The Annals of Statistics* 37, 4254–4278.

LAM, C. AND FENG, P. (2018). A nonparametric eigenvalue-regularized integrated covariance matrix estimator for asset return data. *Journal of Econometrics* 206, 226–257.

LI, Q. AND RACINE, J. (2007). *Nonparametric Econometrics*. Princeton University Press, Princeton.

PARK, S., HONG, S.Y. AND LINTON, O. (2016). Estimating the quadratic covariation matrix for asynchronously observed high frequency stock returns corrupted by additive measurement error. *Journal of Econometrics* 191, 325-347.

PODOLSKIJ, M. AND VETTER, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* 15, 634–658.

REVUZ, D. AND YOR, M. (1999). *Continuous Martingales and Brownian Motion*. Grundlehren der mathematischen Wissenschaften 293, Springer.

SHEPHARD, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.

TAO, M., WANG, Y. AND ZHOU, H. H. (2013). Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *Annals of Statistics* 41, 1816–1864.

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics.

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

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

XIAO, Z. AND LINTON, O. (2002). A nonparametric prewhitened covariance estimator. *Journal of Time Series Analysis* 23, 215–250.

YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research* 11, 2261–2286.

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

ZHANG, L., MYKLAND, P. A. AND AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.

ZHENG, X. AND LI, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Annals of Statistics* 39, 3121–3151.

ZU, Y. AND BOSWIJK, H. P. (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics* 181, 117–135.