Econometric Theory, **42**, 2026, 63–100.
doi:10.1017/S0266466624000264

NONPARAMETRIC ESTIMATION OF LARGE SPOT VOLATILITY MATRICES FOR HIGH-FREQUENCY FINANCIAL DATA

RUIJUN BU
University of Liverpool

DEGUI LI
University of Macau

OLIVER LINTON
University of Cambridge

HANCHAO WANG
Shandong University

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 generalized 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 high-frequency data contaminated by microstructure noise, we introduce a localized pre-averaging estimation method that reduces the effective magnitude of the noise. We then use the estimation tool developed in the noise-free scenario and derive the uniform convergence rates for the developed spot volatility matrix estimator. We further combine kernel smoothing with the shrinkage technique to estimate the time-varying volatility matrix of the high-dimensional noise vector. In addition, we consider large spot volatility matrix estimation in time-varying factor models with observable risk factors and derive the uniform convergence property. We provide numerical studies including simulation and empirical application to examine the performance of the proposed estimation methods in finite samples.

The authors would like to thank the Co-Editor and two reviewers for the constructive comments, which helped to substantially improve the article. The first author's research was partly supported by the BA Talent Development Award (Grant No. TDA21\210027). The second author's research was partly supported by the Leverhulme Research Fellowship (Grant No. RF-2023-396\9), the BA/Leverhulme Small Research Grant (Grant No. SRG1920/100603), and the National Natural Science Foundation of China (Grant No. 72033002). Address correspondence to Oliver Linton, Faculty of Economics, University of Cambridge, Cambridge, UK; e-mail: obl20@cam.ac.uk.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

1. INTRODUCTION

Modeling 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 high-frequency data analysis. One of the main components in these models is the volatility function or matrix. In the low-dimensional setting (with a single or a small number of assets), the realized volatility 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 realized volatility if the noise is ignored. Hence, various modification techniques such as the two-scale, pre-averaging and realized kernel have been introduced to account for the microstructure noise and produce consistent volatility estimation (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 practical applications, 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 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-regularized integrated covariance matrix for high-dimensional asset returns. Given that there often exists co-movement between a large number of assets driven by either observable or latent risk factors, Fan, Furger, and Xiu (2016), Aït-Sahalia and Xiu (2017), and Dai, Lu, and Xiu (2019) extend the methodologies developed by Fan, Liao, and Mincheva (2011, 2013) to estimate the large volatility matrix by imposing a continuous-time factor model structure on the high-dimensional and high-frequency data. Aït-Sahalia and Xiu (2019) study the principal component analysis of high-frequency data and derive the asymptotic distribution for the realized eigenvalues, eigenvectors, and principal components.

The estimation methodologies in the aforementioned literature often rely on the realized volatility (or covariance) matrices, measuring the volatility structure over a fixed time interval. In practice, it is often interesting to further explore the actual

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

spot/instantaneous volatility structure and its dynamic change over time, which is a particularly important measurement for 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 pointwise and global asymptotic distribution theory and uniform consistency. For the noise-contaminated high-frequency data, Zu and Boswijk (2014) combine the two-scale realized 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). It seems straightforward to extend this local nonparametric method to estimate the spot volatility matrix for a fixed number of assets. However, a further extension to the setting with vast financial assets is nontrivial.

The main methodological and theoretical contributions of this paper are summarized as follows.

- *Large spot volatility matrix estimation with noise-free high-frequency data*. We 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 generalized shrinkage to off-diagonal estimated entries. With small off-diagonal entries forced to be zeros, the resulting large spot volatility matrix estimate would be nondegenerate with stable performance in finite samples. We derive the consistency property for the proposed spot volatility matrix estimator uniformly over the entire time interval 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 rate is 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 interval.
- *Large spot volatility matrix estimation with noise-contaminated high-frequency data and time-varying noise volatility matrix estimation*. When the high-frequency data are contaminated by the microstructure noise, we extend Kanaya and Kristensen's (2016) localized pre-averaging estimation method to the high-dimensional setup. Specifically, we first pre-average the noise-contaminated log prices 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. The microstructure noise vector is assumed to be heteroskedastic with the time-varying covariance structure satisfying the uniform sparsity assumption. We show that the existence of microstructure noises slows down the uniform convergence rates (see Theorem 2). Furthermore, we combine the kernel smoothing with generalized shrinkage to estimate the time-varying noise volatility matrix and derive its uniform convergence property. To the best of our

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

knowledge, there is virtually no work on high-dimensional time-varying noise volatility matrix estimation for high-frequency data.

- *Large spot volatility matrix estimation with risk factors*. Since the uniform sparsity assumption is often too restrictive, we relax this restriction in Section 4 and consider large spot volatility matrix estimation in the time-varying factor model at high frequency, i.e., a large number of asset prices are driven by a small number of observable common factors. By imposing the sparsity restriction on the spot idiosyncratic volatility matrix, we obtain the so-called "low-rank plus sparse" spot volatility structure. A similar structure (with constant betas) is adopted by Fan et al. (2016) and Dai et al. (2019) in estimation of large volatility matrices. We use the kernel smoothing method to estimate the spot volatility and covariance of the observed asset prices and factors as well as the time-varying betas, and apply the shrinkage technique to the estimated spot idiosyncratic volatility matrix. We derive the uniform convergence property of the developed matrix estimates, partly extending the pointwise convergence property in Kong (2018). The developed methodology and theory can be further modified to tackle the noise-contaminated high-frequency data.

The rest of the paper is organized 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 high-frequency data setting and further estimate the time-varying noise volatility matrix. Section 4 considers the large spot volatility matrix with systematic factors. Section 5 reports the simulation studies, and Section 6 provides an empirical application. Section 7 concludes the paper and discusses modification of the estimation methodology to allow for jumps in the price or volatility. Proofs of the main theoretical results are available in Appendix A. The Supplementary Material contains proofs of some technical lemmas and propositions and discussions on the spot precision matrix estimation and the asynchronicity issue. 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, respectively, $|\mathbf{A}|_1 = \sum_{i=1}^d \sum_{j=1}^d |A_{ij}|$, $\|\mathbf{A}\|_1 = \max_{1\le j\le d} \sum_{i=1}^d |A_{ij}|$, $\|\mathbf{A}\|_{\infty,q} = \max_{1\le i\le d} \sum_{j=1}^d |A_{ij}|^q$, and $\|\mathbf{A}\|_{\max} = \max_{1\le i\le d} \max_{1\le j\le d} |A_{ij}|$.

2. ESTIMATION WITH NOISE-FREE DATA

Suppose that $\mathbf{X}_t = \left(X_{1,t}, \ldots, X_{p,t}\right)^\intercal$ 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 $\mathbf{W}_t = \left(W_{1,t}, \ldots, W_{p,t}\right)^\intercal$ is a p-dimensional standard Brownian motion, $\boldsymbol{\mu}_t = (\mu_{1,t}, \ldots, \mu_{p,t})^\intercal$ is a p-dimensional drift vector, and $\boldsymbol{\sigma}_t = \left(\sigma_{ij,t}\right)_{p\times p}$ is a $p \times p$

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

matrix. The spot volatility matrix of $\mathbf{X}_t$ is defined as

$$\mathbf{\Sigma}_t = \left(\Sigma_{ij,t}\right)_{p\times p} = \boldsymbol{\sigma}_t \boldsymbol{\sigma}_t^{\mathsf{T}}. \tag{2.2}$$

Our main interest lies in estimating $\mathbf{\Sigma}_t$ when p is large. As in Chen et al. (2013) and Chen and Leng (2016), we assume that the true spot volatility matrix satisfies the following uniform sparsity condition: $\{\mathbf{\Sigma}_t : \ 0 \le t \le T\} \in \mathcal{S}(q, \varpi(p), T)$, where

$$\mathcal{S}(q, \varpi(p), T) = \left\{ \mathbf{\Gamma}_t = \left(\Gamma_{ij,t}\right)_{p\times p},\ t \in [0,T] \ \Big|\ \sup_{0\le t\le T} \|\mathbf{\Gamma}_t\|_{\infty,q} \le \Lambda \varpi(p) \right\}, \tag{2.3}$$

where $0 \le q < 1$, $\varpi(p)$ is larger than a positive constant, T is a fixed positive number, and Λ is a positive random variable satisfying $\mathsf{E}[\Lambda] \le C_\Lambda < \infty$. This is a natural extension of the approximate sparsity assumption (e.g., Bickel and Levina, 2008; Wang and Zou, 2010; Tao et al., 2013). Section 4 below will relax this assumption and consider estimating large spot volatility matrices with systematic factors. The asset prices are assumed to be collected over a fixed time interval $[0,T]$ at $0, \Delta, 2\Delta, \ldots, n\Delta$, where Δ is the sampling interval and $n = \lfloor T/\Delta \rfloor$ with $\lfloor \cdot \rfloor$ denoting the floor function. In the main text, we focus on the case of equidistant time points in the high-frequency data collection. The asynchronicity issue will be discussed in Appendix C.2 of the Supplementary Material.

For each $1 \le i,j \le p$, we estimate the spot covariance $\Sigma_{ij,t}$ by

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

with

$$K_h^*(t_k - t) = K_h(t_k - t) \Big/ \left[\Delta \sum_{l=1}^{n} K_h(t_l - t) \right],$$

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 $\Delta X_{i,k} = X_{i,t_k} - X_{i,t_{k-1}}$. The use of $K_h^*(t_k - t)$ rather than $K_h(t_k - t)$ in (2.4) is to correct a constant bias when t is close to the boundary points 0 and T. A naive method of estimating the spot volatility matrix $\mathbf{\Sigma}_t$ is to directly use $\widehat{\Sigma}_{ij,t}$ to form an estimated matrix. 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 $i \ne j$, forcing very small estimated off-diagonal entries to be zeros. Let $s_\rho(\cdot)$ denote a shrinkage function satisfying the following three conditions: (i) $|s_\rho(u)| \le |u|$ for $u \in \mathscr{R}$; (ii) $s_\rho(u) = 0$ if $|u| \le \rho$; and (iii) $|s_\rho(u) - u| \le \rho$, where ρ is a user-specified tuning parameter. With the shrinkage function, we construct the following nonparametric estimator of $\mathbf{\Sigma}_t$:

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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

where $\rho_1(t)$ is a tuning parameter which is allowed to change over t and $I(\cdot)$ denotes the indicator function. Section 5 discusses the choice of $\rho_1(t)$, ensuring that $\widehat{\mathbf{\Sigma}}_t$ is positive-definite. Our estimation method of the spot volatility matrix can be seen as a natural extension of the kernel-based large sparse covariance matrix estimation (e.g., Chen et al., 2013; Chen and Leng, 2016; Chen et al., 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,

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

where $\Sigma^*_{ij,s}=\Sigma_{ii,s}+\Sigma_{jj,s}+2\Sigma_{ij,s}$. For the spot covariance process $\{\Sigma_{ij,t}\}$, there exist $\gamma\in(0,1)$ and $B(t,\epsilon)$, a positive random function slowly varying at $\epsilon=0$ and continuous with respect to t, such that

$$\max_{1\le i,j\le p}\left|\Sigma_{ij,t+\epsilon}-\Sigma_{ij,t}\right|\le B(t,\epsilon)|\epsilon|^{\gamma}+o(|\epsilon|^{\gamma}),\ \ \epsilon\to 0. \tag{2.6}$$

Assumption 2.

(i) The kernel $K(\cdot)$ is a bounded and Lipschitz continuous function with a compact support $[-1,1]$. In addition, $\int_{-1}^{1}K(u)du=1$.

(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 generalized shrinkage be chosen as

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

where γ is defined in (2.6) and $M(t)$ is a positive function satisfying that

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

Remark 1. (i) Assumption 1 imposes some mild restrictions on the drift and volatility processes. By a typical localization 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 uniform boundedness over the entire time interval, i.e., with probability one,

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

which are the same as Assumption A2 in Tao et al. (2013) and Assumptions (A.ii) and (A.iii) in Cai et al. (2020). It may be possible to relax the uniform boundedness restriction (when T is allowed to diverge) at the cost of more lengthy

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

proofs (e.g., Kanaya and Kristensen, 2016). Assumption 1(ii) gives the smoothness condition on the spot covariance process, which is crucial to the derivation of the uniform asymptotic order for the kernel estimation bias. In the low-dimensional setting when p is fixed, the smoothness condition (2.6) is standard. For example, when the spot covariance is driven by continuous semimartingales, (2.6) holds with $\gamma < 1/2$ (e.g., Ch. V, Exercise 1.20 in Revuz and Yor, 1999). In the high-dimensional setting, as we need to derive the uniform estimation bias rate, we require the smoothness condition uniformly over the index pairs (i,j). If $\Sigma_{ij,t}$ is a deterministic function of t as in Kristensen (2010) and Remark 2(ii), the uniform smoothness condition (2.6) would be comparable to those in Chen and Leng (2016) and Chen et al. (2019). However, the smoothness restriction rules out volatility jumps, in which case the developed estimation methodology needs to be modified (see the discussion in Section 7).

(ii) Assumption 2(i) contains some commonly used conditions for the kernel function. Assumption 2(ii) and (iii) imposes 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 conventional bandwidth restriction. Assumption 2(iii) is comparable to that assumed by Chen and Leng (2016) and Chen et al. (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 entries in the spot volatility matrix estimation. For example, we set $\rho_{ij}(t) = \rho(t)(\widehat{\Sigma}_{ii,t}\widehat{\Sigma}_{jj,t})^{1/2}$ in the numerical studies and shrink the (i,j)-entry to zero if $\widehat{\Sigma}_{ij,t} \le \rho(t)(\widehat{\Sigma}_{ii,t}\widehat{\Sigma}_{jj,t})^{1/2}$.

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

THEOREM 1. *Suppose that Assumptions 1 and 2 are satisfied, and* $\{\mathbf{\Sigma}_t : 0 \le t \le T\} \in \mathcal{S}(q, \varpi(p), T)$. *Then we have*

$$\sup_{0\le t\le T} \left\| \widehat{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t \right\| = O_P\left(\varpi(p) \zeta_{\Delta,p}^{1-q} \right), \tag{2.7}$$

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

Remark 2. (i) The first term of $\zeta_{\Delta,p}$ is h^γ, which is the bias rate due to application of the local smoothing technique. It is slower than the conventional h^2-rate since we do not assume existence of smooth derivatives of $\Sigma_{ij,t}$ (with respect to t). The second term of $\zeta_{\Delta,p}$ is the square root of $\Delta h^{-1}\log(p \vee \Delta^{-1})$, a typical uniform asymptotic rate for the kernel estimation variance component. The uniform convergence rate in (2.7) is also similar to those obtained by Chen and Leng (2016) and Chen et al. (2019) in the low-frequency data setting (disregarding the bias order). 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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

ultrahigh-dimensional setting when p diverges at an exponential rate of $n = \lfloor T/\Delta \rfloor$. Treating (nh) as the "effective" sample size in the local estimation procedure and disregarding the bias rate h^γ, the rate in (2.7) is comparable to the optimal minimax rate in large covariance matrix estimation (e.g., Cai and Zhou, 2012).

(ii) If we further assume that $\Sigma_{ij,t}$ is deterministic with continuous second-order derivative with respect to t, and $K(\cdot)$ is symmetric, we may improve the kernel estimation bias order. In fact, following the proof of Theorem 1, we may show that

$$\sup_{h \le t \le T-h} \left\| \widehat{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t \right\| = O_P\left(\varpi(p) \zeta_{\Delta,p,\star}^{1-q} \right), \tag{2.8}$$

where $\zeta_{\Delta,p,\star} = h^2 + \left[\frac{\Delta \log(p \vee \Delta^{-1})}{h} \right]^{1/2}$. The above uniform consistency property only holds over the trimmed time interval $[h, T-h]$ due to the kernel boundary effect. In practice, however, 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. To address this issue, we may replace $K_h^*(t_k - t)$ in (2.4) by a boundary kernel weight defined by

$$K_{h,t}^{\star}(t_k - t) = K_t\left(\frac{t_k - t}{h} \right) \Big/ \left[\Delta \sum_{l=1}^{n} K_t\left(\frac{t_l - t}{h} \right) \right],$$

where $K_t(\cdot)$ is a boundary kernel satisfying $\int_{-t/h}^{(T-t)/h} u K_t(u) du = 0$ (a key condition to improve the bias order near the boundary points). Examples of boundary kernels can be found in Fan and Gijbels (1996) and Li and Racine (2007). With this adjustment in the kernel estimation, we can extend the uniform consistency result (2.8) to the entire interval $[0, T]$.

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. 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 $t_k = k\Delta$, $k = 1, \ldots, n$, $\mathbf{Z}_t = \left(Z_{1,t}, \ldots, Z_{p,t}\right)^\intercal$ is a vector of observed asset prices at time t, and $\boldsymbol{\xi}_k = (\xi_{1,k}, \ldots, \xi_{p,k})^\intercal$ is a p-dimensional vector of noises with nonlinear heteroskedasticity, $\boldsymbol{\omega}(\cdot) = \left[\omega_{ij}(\cdot)\right]_{p \times p}$ is a $p \times p$ matrix of deterministic functions, and $\boldsymbol{\xi}_k^* = \left(\xi_{1,k}^*, \ldots, \xi_{p,k}^*\right)^\intercal$ independently follows a p-variate identical distribution with identity covariance matrix. The noise structure defined in (3.1) is similar to Kalnina and Linton's (2008) setting which also contains a nonlin-

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

ear 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 the pre-averaging technique as the realized 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 et al., 2005) can be re-written as the realized 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 et al. (2010) in estimating the volatility for a single asset and is further extended by Kim et al. (2016) and Dai et al. (2019) to the large high-frequency data setting. Kanaya and Kristensen (2016) use a localized 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.

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

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

with $L_b^\dagger(t_k-\tau) = L_b(t_k-\tau)/\int_0^T L_b(s-\tau)ds$, where $L_b(u) = b^{-1}L(u/b)$, $L(\cdot)$ 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 ith component of $\widetilde{\mathbf{X}}_{\tau_l}$ and $\tau_0, \tau_1, \ldots, \tau_N$ 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), we estimate the spot covariance $\Sigma_{ij,t}$ by

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

where

$$K_h^\dagger(\tau_l - t) = K_h(\tau_l - t)\Big/\left[\Delta_*\sum_{k=1}^{N} K_h(\tau_k - t)\right].$$

Furthermore, to obtain a stable spot volatility matrix estimate in finite samples when p is large, as in (2.5), we apply shrinkage to $\widetilde{\Sigma}_{ij,t}$, $1 \le i \ne j \le p$, and subsequently construct

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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

where $\rho_2(t)$ is a 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 $\mathsf{E}(\xi^*_{i,k}) = 0$ and

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

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

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

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

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 $0 < s \le s_0$.

(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 moment condition in Assumption 3(i) is weaker than the sub-Gaussian condition (e.g., Bickel and Levina, 2008; Tao et al., 2013), which is commonly used in large covariance matrix estimation when p is ultra large. The boundedness condition on $\omega_{ij}(\cdot)$ in Assumption 3(ii) is comparable to the local boundedness restriction in Assumption 1(i). Assumption 4(ii) imposes some mild restrictions on b and p, which 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 $\widetilde{\boldsymbol{X}}_\tau$ 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).

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

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

$$\sup_{0\le t\le T} \left\| \widetilde{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t \right\| = O_P\left(\varpi(p) \left[\zeta^*_{N,p} + \nu_{\Delta,p,N} \right]^{1-q} \right), \tag{3.5}$$

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 interval $\Delta_* = T/N$, 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 $\widetilde{\boldsymbol{X}}_\tau$ in the first step of pre-averaging estimation. In particular, when $q = 0$, $\varpi(p)$ is bounded, $b = \Delta^{1/4}$ and $h = N^{-\frac{1}{2\gamma+1}}$ with $N = \Delta^{-\frac{2\gamma+1}{2(4\gamma+1)}}$, the uniform convergence rate in (3.5) becomes $\Delta^{\frac{\gamma}{2(4\gamma+1)}}\sqrt{\log(p \vee \Delta^{-1})}$. Furthermore, if $\gamma = 1/2$, 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.

3.2. Estimation of the Time-Varying Noise Volatility Matrix

It is often interesting to further explore the volatility structure of microstructure noise. Chang et al. (2024) estimate the constant covariance matrix for high-dimensional noise and derive the optimal convergence rates for the developed estimate. In the present paper, we consider the time-varying noise covariance matrix defined by

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

It is sensible to assume that $\{\mathbf{\Omega}(t) : 0 \le t \le T\}$ satisfies the uniform sparsity condition as in (2.3). For each $1 \le i,j \le p$, 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, $\Delta Z_{i,t_k} = Z_{i,t_k} - Z_{i,t_{k-1}}$ and $K^*_{h_1}(t_k - t)$ is defined similarly to $K^*_h(t_k - t)$ in (2.4) but with h_1 replacing h. As in (2.5) and (3.4), we again apply shrinkage to $\widehat{\Omega}_{ij}(t)$, $1 \le i \ne j \le p$, and construct

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

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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

Assumption 5.

(i) For any p-dimensional vector $\mathbf{u}$ satisfying $\|\mathbf{u}\|_2 = 1$, $\mathsf{E}\left[\exp\left(s(\mathbf{u}^\mathsf{T}\boldsymbol{\xi}_k^*)^2\right)\right] \le C_\xi^\star < \infty$, $0 < s \le s_0$.

(ii) The time-varying function $\Omega_{ij}(t)$ satisfies that

$$\max_{1\le i,j\le p} \left|\Omega_{ij}(t) - \Omega_{ij}(s)\right| \le C_\Omega |t-s|^{\gamma_1},$$

where C_Ω is a positive constant and $0 < \gamma_1 < 1$.

(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 $0 < \iota_\star < 1/2$, $0 < s \le s_0$ 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 et al., 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 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, and Assumption 2(ii) and (iii) holds when $\rho_1(t)$, $\zeta_{\Delta,p}$, and h are replaced by $\rho_3(t)$, $\delta_{\Delta,p}$, and h_1, respectively, where $\delta_{\Delta,p} = h_1^{\gamma_1} + \left[\frac{\Delta\log(p\vee\Delta^{-1})}{h_1}\right]^{1/2}$. If $\{\mathbf{\Omega}(t) : 0 \le t \le T\} \in \mathcal{S}(q, \varpi(p), T)$, we have*

$$\sup_{0\le t\le T} \left\|\widehat{\mathbf{\Omega}}(t) - \mathbf{\Omega}(t)\right\| = O_P\left(\varpi(p)\delta_{\Delta,p}^{1-q}\right). \tag{3.9}$$

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. (2024) for the constant noise covariance matrix estimation. Meanwhile, the kernel estimation bias order $h_1^{\gamma_1}$ may be improved by strengthening the smoothness condition on $\Omega_{ij}(\cdot)$ and adopting the boundary kernel weight as suggested in Remark 2(ii). It is worth pointing out that the estimation in (3.7) relies on the independence assumption on the microstructure noise. It becomes inconsistent when the noise is temporally correlated and an appropriate modification technique such as ReMeDI in Li and Linton (2022) may be required.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

4. ESTIMATION WITH OBSERVED FACTORS

The large spot volatility matrix estimation with the shrinkage technique developed in Sections 2 and 3 heavily relies on the uniform sparsity assumption (2.3). However, the latter may be too restrictive in practice since the price processes of a large number of assets are often driven by some common factors such as the market factor, resulting in strong correlation among assets and failure of the sparsity condition. To address this problem, we next consider the nonparametric time-varying regression at high frequency:

$$d\mathbf{Y}_t = \boldsymbol{\beta}(t)d\mathbf{F}_t + d\mathbf{X}_t, \tag{4.1}$$

where $\boldsymbol{\beta}(t) = \left[\beta_1(t), \ldots, \beta_p(t)\right]^\intercal$ is a $p \times k$ matrix of time-varying betas (or factor loadings), $\mathbf{F}_t$ and $\mathbf{X}_t$ are k-variate and p-variate continuous semi-martingales defined by

$$d\mathbf{F}_t = \boldsymbol{\mu}_t^F dt + \boldsymbol{\sigma}_t^F d\mathbf{W}_t^F \text{ and } d\mathbf{X}_t = \boldsymbol{\mu}_t^X dt + \boldsymbol{\sigma}_t^X d\mathbf{W}_t^X, \tag{4.2}$$

respectively, $\boldsymbol{\mu}_t^F$ and $\boldsymbol{\mu}_t^X$ are drift vectors, $\boldsymbol{\sigma}_t^F = \left(\sigma_{ij,t}^F\right)_{k\times k}$, $\boldsymbol{\sigma}_t^X = \left(\sigma_{ij,t}^X\right)_{p\times p}$, $\mathbf{W}_t^F$ and $\mathbf{W}_t^X$ are k-dimensional and p-dimensional standard Brownian motions. For the time being, we assume that $\mathbf{Y}_t$ and $\mathbf{F}_t$ are observable and noise free but $\mathbf{X}_t$ is latent. Extension of the methodology and theory to the noise-contaminated high-frequency data will be considered later in this section.

Estimation of the constant betas via the ratio of realized covariance to realized variance is proposed by Barndorff-Nielsen and Shephard (2004), and extension to time-varying beta estimation has been studied by Mykland and Zhang (2006), Reiß, Todorov, and Tauchen (2015), Aït-Sahalia, Kalnina, and Xiu (2020), and Andersen, Thyrsgaard, and Todorov (2021), some of which allow for jumps in the semimartingale processes. For example, Andersen et al. (2021) consider the intraday variation in the cross-sectional dispersion of time-varying betas when the asset number is either fixed or divergent; Liao and Todorov (2024) test changes in the span of betas with latent systematic risk factors. The main interest of this section lies in estimating the large spot volatility structure $\boldsymbol{\Sigma}_t^Y$ of $\mathbf{Y}_t$ within the model framework (4.1). Letting $\boldsymbol{\Sigma}_t^F = \boldsymbol{\sigma}_t^F \left(\boldsymbol{\sigma}_t^F\right)^\intercal$ and $\boldsymbol{\Sigma}_t^X = \boldsymbol{\sigma}_t^X \left(\boldsymbol{\sigma}_t^X\right)^\intercal$ and assuming orthogonality between $\mathbf{X}_t$ and $\mathbf{F}_t$ (see Assumption 6(iii) below), it follows from (4.1) that

$$\boldsymbol{\Sigma}_t^Y = \boldsymbol{\beta}(t)\boldsymbol{\Sigma}_t^F \boldsymbol{\beta}(t)^\intercal + \boldsymbol{\Sigma}_t^X. \tag{4.3}$$

As in Fan et al. (2011, 2013), we impose the uniform sparsity restriction on $\boldsymbol{\Sigma}_t^X$ instead of $\boldsymbol{\Sigma}_t^Y$, i.e., $\left\{\boldsymbol{\Sigma}_t^X : \ 0 \le t \le T\right\} \in \mathcal{S}(q, \varpi(p), T)$. This is a reasonable assumption in practical applications as the asset prices, after removing the influence of systematic factors, are expected to be weakly correlated. Fan et al. (2016) and Dai et al. (2019) use a similar framework with constant betas to estimate large volatility matrices.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

Suppose that we observe $\mathbf{Y}_t$ and $\mathbf{F}_t$ at regular points: $t_k = k\Delta$, $k = 1, \ldots, n$, as in Sections 2 and 3. Let $\mathbf{\Sigma}_t^{YF}$ be the spot covariance between $\mathbf{Y}_t$ and $\mathbf{F}_t$. We may use the kernel smoothing method as in (2.4) to estimate $\mathbf{\Sigma}_t^Y$, $\mathbf{\Sigma}_t^F$ and $\mathbf{\Sigma}_t^{YF}$, i.e.,

$$\widehat{\mathbf{\Sigma}}_t^Y = \sum_{k=1}^{n} K_h^*(t_k - t)\Delta\mathbf{Y}_k\Delta\mathbf{Y}_k^\intercal, \tag{4.4}$$

$$\widehat{\mathbf{\Sigma}}_t^F = \sum_{k=1}^{n} K_h^*(t_k - t)\Delta\mathbf{F}_k\Delta\mathbf{F}_k^\intercal, \tag{4.5}$$

$$\widehat{\mathbf{\Sigma}}_t^{YF} = \sum_{k=1}^{n} K_h^*(t_k - t)\Delta\mathbf{Y}_k\Delta\mathbf{F}_k^\intercal, \tag{4.6}$$

where $\Delta\mathbf{Y}_k = \mathbf{Y}_{t_k} - \mathbf{Y}_{t_{k-1}}$, $\Delta\mathbf{F}_k = \mathbf{F}_{t_k} - \mathbf{F}_{t_{k-1}}$, and $K_h^*(t_k - t)$ is defined as in (2.4). Consequently, the time-varying betas $\boldsymbol{\beta}(t)$ and the spot idiosyncratic volatility matrix $\mathbf{\Sigma}_t^X$ are estimated by

$$\widehat{\boldsymbol{\beta}}(t) = \left[\widehat{\beta}_1(t), \ldots, \widehat{\beta}_p(t)\right]^\intercal = \widehat{\mathbf{\Sigma}}_t^{YF}\left(\widehat{\mathbf{\Sigma}}_t^F\right)^{-1} \tag{4.7}$$

and

$$\widehat{\mathbf{\Sigma}}_t^X = \left(\widehat{\Sigma}_{ij,t}^X\right)_{p\times p} = \widehat{\mathbf{\Sigma}}_t^Y - \widehat{\mathbf{\Sigma}}_t^{YF}\left(\widehat{\mathbf{\Sigma}}_t^F\right)^{-1}\left(\widehat{\mathbf{\Sigma}}_t^{YF}\right)^\intercal. \tag{4.8}$$

With the uniform sparsity condition, it is sensible to further apply shrinkage to $\widehat{\Sigma}_{ij,t}^X$, i.e.,

$$\widehat{\mathbf{\Sigma}}_t^{X,s} = \left(\widehat{\Sigma}_{ij,t}^{X,s}\right)_{p\times p} \quad \text{with} \quad \widehat{\Sigma}_{ij,t}^{X,s} = s_{\rho_4(t)}(\widehat{\Sigma}_{ij,t}^X)I(i \neq j) + \widehat{\Sigma}_{ii,t}^X I(i = j), \tag{4.9}$$

where $\rho_4(t)$ is a time-varying shrinkage parameter. We finally estimate $\mathbf{\Sigma}_t^Y$ as

$$\widehat{\mathbf{\Sigma}}_t^{Y,s} = \widehat{\boldsymbol{\beta}}(t)\widehat{\mathbf{\Sigma}}_t^F\widehat{\boldsymbol{\beta}}(t)^\intercal + \widehat{\mathbf{\Sigma}}_t^{X,s} = \widehat{\mathbf{\Sigma}}_t^{YF}\left(\widehat{\mathbf{\Sigma}}_t^F\right)^{-1}\left(\widehat{\mathbf{\Sigma}}_t^{YF}\right)^\intercal + \widehat{\mathbf{\Sigma}}_t^{X,s}. \tag{4.10}$$

We need the following assumption to derive the uniform convergence property for $\widehat{\mathbf{\Sigma}}_t^{X,s}$ and $\widehat{\mathbf{\Sigma}}_t^{Y,s}$.

Assumption 6.

(i) Assumption 1 is satisfied for $\{X_t\}$ defined in (4.2) (with minor notational changes).
(ii) Let $\{\boldsymbol{\mu}_t^F\}$, $\{\boldsymbol{\sigma}_t^F\}$, and $\{\mathbf{\Sigma}_t^F\}$ satisfy the boundedness and smoothing conditions as in Assumption 1.
(iii) For any $1 \le i \le p$ and $1 \le j \le k$, $\left[X_{it}, F_{jt}\right] = 0$ for any $t \in [0, T]$, where $X_{i,t}$ is the ith element of $\mathbf{X}_t$, $F_{j,t}$ is the j-th element of $\mathbf{F}_t$, and $[\cdot,\cdot]$ denotes the quadratic covariation.
(iv) The time-varying beta function $\beta_i(\cdot)$ satisfies that

$$\max_{1\le i\le p}\sup_{0\le t\le T}\|\beta_i(t)\|_2 \le C_\beta < \infty, \quad \max_{1\le i\le p}\|\beta_i(t) - \beta_i(s)\|_2 \le C_\beta|t - s|^\gamma,$$

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

where γ is the same as that in Assumption 1(ii). In addition, there exists a positive definite matrix $\mathbf{\Sigma}_\beta(t)$ (with uniformly bounded eigenvalues) such that

$$\sup_{0\le t\le T}\left\|\frac{1}{p}\boldsymbol{\beta}(t)^\intercal\boldsymbol{\beta}(t)-\mathbf{\Sigma}_\beta(t)\right\|=o(1). \tag{4.11}$$

Remark 7. The uniform boundedness and smoothness conditions imposed on the drift and spot volatility functions of $\mathbf{X}_t$ and $\mathbf{F}_t$ in Assumption 6(i) and (ii) are the same as those in Assumption 1. This is crucial to ensure that the uniform convergence rates of $\widehat{\mathbf{\Sigma}}_t^Y$, $\widehat{\mathbf{\Sigma}}_t^F$, and $\widehat{\mathbf{\Sigma}}_t^{YF}$ (in the max norm) derived in Proposition A.4 are the same as that in Proposition A.1. The orthogonality condition in Assumption 6(iii) is commonly used to consistently estimate the time-varying factor model (e.g., Fan et al., 2016; Dai et al., 2019). Assumption 6(iv) is a rather mild restriction on time-varying betas and may be strengthened to improve the estimation bias order (see the discussion in Remark 2(ii)). The condition (4.11) indicates that all the factors are pervasive.

We next present the convergence property of $\widehat{\mathbf{\Sigma}}_t^{X,s}$ and $\widehat{\mathbf{\Sigma}}_t^{Y,s}$ defined in (4.9) and (4.10), respectively. Due to the nonparametric factor regression model structure (4.1), the largest k eigenvalues of $\mathbf{\Sigma}_t^Y$ are spiked, diverging at a rate of p. Hence, $\mathbf{\Sigma}_t^Y$ cannot be consistently estimated in the absolute term. To address this problem, as in Fan et al. (2011, 2013), we measure the spiked volatility matrix estimate in the following relative error:

$$\left\|\widehat{\mathbf{\Sigma}}_t^{Y,s}-\mathbf{\Sigma}_t^Y\right\|_{\mathbf{\Sigma}_t^Y}=\frac{1}{\sqrt{p}}\left\|\left(\mathbf{\Sigma}_t^Y\right)^{-1/2}\left(\widehat{\mathbf{\Sigma}}_t^{Y,s}-\mathbf{\Sigma}_t^Y\right)\left(\mathbf{\Sigma}_t^Y\right)^{-1/2}\right\|_F,$$

where the normalization factor $p^{-1/2}$ is used to guarantee that $\left\|\mathbf{\Sigma}_t^Y\right\|_{\mathbf{\Sigma}_t^Y}=1$.

THEOREM 4. *Suppose that Assumptions 2(i) and (ii) and 6 are satisfied, and Assumption 2(iii) holds with $\rho_1(t)$ replaced by $\rho_4(t)$. When $\left\{\mathbf{\Sigma}_t^X:\ 0\le t\le T\right\}\in\mathcal{S}(q,\varpi(p),T)$, we have*

$$\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}_t^{X,s}-\mathbf{\Sigma}_t^X\right\|=O_P\left(\varpi(p)\zeta_{\Delta,p}^{1-q}\right), \tag{4.12}$$

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

$$\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}_t^{Y,s}-\mathbf{\Sigma}_t^Y\right\|_{\mathbf{\Sigma}_t^Y}=O_P\left(p^{1/2}\zeta_{\Delta,p}^2+\varpi(p)\zeta_{\Delta,p}^{1-q}\right). \tag{4.13}$$

Remark 8. Although $\mathbf{X}_t$ is latent in model (4.1), the uniform convergence rate for $\widehat{\mathbf{\Sigma}}_t^{X,s}$ in (4.12) is the same as that in Theorem 1 when $\mathbf{X}_t$ is observable. Treating (nh) as the effective sample size in kernel estimation and disregarding the bias order in $\zeta_{\Delta,p}$, the uniform convergence rate for $\widehat{\mathbf{\Sigma}}_t^{Y,s}$ in (4.13) is comparable to the convergence rates derived by Fan et al. (2011) in low frequency and Fan et al. (2016) in high frequency. To guarantee uniform consistency in the relative

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

matrix estimation error, we have to further assume that $p\zeta_{\Delta,p}^4 = o(1)$, limiting the divergence rate of the asset number, i.e., p can only diverge at a polynomial rate of $n = \lfloor T/\Delta \rfloor$.

We next modify the above methodology and theory to accommodate microstructure noise in the asset prices and factors. Assume that

$$\mathbf{Z}_{Y,t_k} = \mathbf{Y}_{t_k} + \boldsymbol{\omega}_Y(t_k)\boldsymbol{\xi}_{Y,k}^*, \quad \mathbf{Z}_{F,t_k} = \mathbf{F}_{t_k} + \boldsymbol{\omega}_F(t_k)\boldsymbol{\xi}_{F,k}^*, \tag{4.14}$$

where $\boldsymbol{\omega}_Y(\cdot)$ and $\boldsymbol{\omega}_F(\cdot)$ are matrices of deterministic functions similar to $\boldsymbol{\omega}(\cdot)$, and $\{\boldsymbol{\xi}_{Y,k}^*\}$ and $\{\boldsymbol{\xi}_{F,k}^*\}$ are i.i.d. sequences of random vectors similar to $\{\boldsymbol{\xi}_k^*\}$. Since both $\mathbf{Y}_t$ and $\mathbf{F}_t$ are latent, we need to first adopt the pre-averaging technique proposed in Section 3.1 to obtain the approximation of $\mathbf{Y}_t$ and $\mathbf{F}_t$, and then apply the kernel smoothing and generalized shrinkage as in (4.4)–(4.10). This results in a three-stage estimation procedure which we describe as follows.

1. As in (3.2), we pre-average the noise-contaminated $\mathbf{Z}_{Y,t_k}$ and $\mathbf{Z}_{F,t_k}$ via the kernel filter:

$$\widetilde{\mathbf{Y}}_\tau = \frac{T}{n}\sum_{k=1}^{n} L_b^\dagger(t_k - \tau)\mathbf{Z}_{Y,t_k}, \quad \widetilde{\mathbf{F}}_\tau = \frac{T}{n}\sum_{k=1}^{n} L_b^\dagger(t_k - \tau)\mathbf{Z}_{F,t_k}, \tag{4.15}$$

   where $L_b^\dagger(t_k - \tau)$ is defined as in (3.2) and we consider τ as the pseudo-sampling time points: $\tau_l = l\Delta_*$, $l = 0, 1, \ldots, N = \lfloor T/\Delta_* \rfloor$.
2. With $\widetilde{\mathbf{Y}}_{\tau_l}$ and $\widetilde{\mathbf{F}}_{\tau_l}$, $l = 1, \ldots, N$, we estimate $\boldsymbol{\Sigma}_t^Y, \boldsymbol{\Sigma}_t^F$, and $\boldsymbol{\Sigma}_t^{YF}$ by the kernel smoothing as in (4.4)–(4.6):

$$\widetilde{\boldsymbol{\Sigma}}_t^Y = \sum_{l=1}^{N} K_h^\dagger(\tau_l - t)\Delta\widetilde{\mathbf{Y}}_l \Delta\widetilde{\mathbf{Y}}_l^\intercal,$$

$$\widetilde{\boldsymbol{\Sigma}}_t^F = \sum_{l=1}^{N} K_h^\dagger(\tau_l - t)\Delta\widetilde{\mathbf{F}}_l \Delta\widetilde{\mathbf{F}}_l^\intercal,$$

$$\widetilde{\boldsymbol{\Sigma}}_t^{YF} = \sum_{l=1}^{N} K_h^\dagger(\tau_l - t)\Delta\widetilde{\mathbf{Y}}_l \Delta\widetilde{\mathbf{F}}_l^\intercal,$$

   where $K_h^\dagger(\tau_l - t)$ is defined as in (3.3), $\Delta\widetilde{\mathbf{Y}}_l = \widetilde{\mathbf{Y}}_{\tau_l} - \widetilde{\mathbf{Y}}_{\tau_{l-1}}$, and $\Delta\widetilde{\mathbf{F}}_l = \widetilde{\mathbf{F}}_{\tau_l} - \widetilde{\mathbf{F}}_{\tau_{l-1}}$. Furthermore, estimate $\boldsymbol{\beta}(t)$ and $\boldsymbol{\Sigma}_t^X$ by

$$\widetilde{\boldsymbol{\beta}}(t) = \widetilde{\boldsymbol{\Sigma}}_t^{YF}\left(\widetilde{\boldsymbol{\Sigma}}_t^F\right)^{-1}, \quad \widetilde{\boldsymbol{\Sigma}}_t^X = \left(\widetilde{\Sigma}_{ij,t}^X\right)_{p\times p} = \widetilde{\boldsymbol{\Sigma}}_t^Y - \widetilde{\boldsymbol{\Sigma}}_t^{YF}\left(\widetilde{\boldsymbol{\Sigma}}_t^F\right)^{-1}\left(\widetilde{\boldsymbol{\Sigma}}_t^{YF}\right)^\intercal.$$

3. Apply the generalized shrinkage to $\widetilde{\Sigma}_{ij,t}^X$, i.e.,

$$\widetilde{\boldsymbol{\Sigma}}_t^{X,s} = \left(\widetilde{\Sigma}_{ij,t}^{X,s}\right)_{p\times p} \text{ with } \widetilde{\Sigma}_{ij,t}^{X,s} = s_{\rho_5(t)}(\widetilde{\Sigma}_{ij,t}^X)I(i \neq j) + \widetilde{\Sigma}_{ii,t}^X I(i = j),$$

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

where $\rho_5(t)$ is the shrinkage parameter, and then estimate $\mathbf{\Sigma}_t^Y$ by

$$\widetilde{\mathbf{\Sigma}}_t^{Y,s} = \widetilde{\boldsymbol{\beta}}(t)\widetilde{\mathbf{\Sigma}}_t^F\widetilde{\boldsymbol{\beta}}(t)^\intercal + \widetilde{\mathbf{\Sigma}}_t^{X,s} = \widetilde{\mathbf{\Sigma}}_t^{YF}\left(\widetilde{\mathbf{\Sigma}}_t^F\right)^{-1}\left(\widetilde{\mathbf{\Sigma}}_t^{YF}\right)^\intercal + \widetilde{\mathbf{\Sigma}}_t^{X,s}.$$

As shown in Theorem 2, the existence of microstructure noises slows down the uniform convergence rates. Following the proof of Lemma B.1 in Appendix B, we may show that

$$\max_{0\le l\le N}\left|\widetilde{\mathbf{Y}}_{\tau_l} - \mathbf{Y}_{\tau_l}\right|_{\max} + \max_{0\le l\le N}\left|\widetilde{\mathbf{F}}_{\tau_l} - \mathbf{F}_{\tau_l}\right|_{\max} = O_P\left(\nu_{\Delta,p,N}\right),$$

where $|\cdot|_{\max}$ denotes the L_∞-norm of a vector, and $\nu_{\Delta,p,N}$ is defined in Assumption 4(iii). Modifying Proposition A.4 and the proof of Theorem 4 in Appendix A, we can prove that (4.12) and (4.13) hold but with $\zeta_{\Delta,p}$ replaced by $\zeta^*_{N,p} + \nu_{\Delta,p,N}$ defined in Assumption 4(iii), i.e.,

$$\sup_{0\le t\le T}\left\|\widetilde{\mathbf{\Sigma}}_t^{X,s} - \mathbf{\Sigma}_t^X\right\| = O_P\left(\varpi(p)(\zeta^*_{N,p} + \nu_{\Delta,p,N})^{1-q}\right),$$

$$\sup_{0\le t\le T}\left\|\widetilde{\mathbf{\Sigma}}_t^{Y,s} - \mathbf{\Sigma}_t^Y\right\|_{\mathbf{\Sigma}_t^Y} = O_P\left(p^{1/2}(\zeta^*_{N,p} + \nu_{\Delta,p,N})^2 + \varpi(p)(\zeta^*_{N,p} + \nu_{\Delta,p,N})^{1-q}\right).$$

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 under the sparsity condition and the factor-based spot volatility matrix estimation. Here we only consider the synchronous high-frequency data. Additional simulation results for asynchronous high-frequency data are provided in the Supplementary Material.

5.1. Simulation for Sparse Volatility Matrix Estimation

5.1.1. *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 $\mathbf{\Omega}(t) = \left[\Omega_{ij}(t)\right]_{p\times p}$, $\boldsymbol{\xi}_k^* = \left(\xi_{1,k}^*, \cdots, \xi_{p,k}^*\right)^\intercal$ is an independent p-dimensional vector of cross-sectionally independent standard normal random variables, the latent return process $\mathbf{X}_t$ of p assets is generated from the following drift-free model:

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

$\mathbf{W}_t^X = \left(W_{1,t}^X, \ldots, W_{p,t}^X\right)^\intercal$ is a standard p-dimensional Brownian motion, and $\boldsymbol{\sigma}_t$ is chosen as the Cholesky decomposition of the spot covariance matrix $\mathbf{\Sigma}_t = \left(\Sigma_{ij,t}\right)_{p\times p}$. 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,

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

i.e., $\Delta = 1/(252 \times 6.5 \times 60 \times 4)$, to generate synchronous data. We consider three structures in $\mathbf{\Sigma}_t$ and $\mathbf{\Omega}(t)$: "banding," "block-diagonal," and "exponentially decaying." Following Wang and Zou (2010), we generate the diagonal elements of $\mathbf{\Sigma}_t$ from the following geometric Ornstein–Uhlenbeck model (e.g., Barndorff-Nielsen and Shephard, 2002):

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

where $\mathbf{W}_t^* = \left(W_{1,t}^*, \ldots, W_{p,t}^*\right)^{\intercal}$ is a p-dimensional standard 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 $\mathbf{\Omega}(t)$ are defined as daily cyclical deterministic functions of time:

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

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

$$\kappa_t^{\Sigma} = \frac{e^{2\kappa_t}-1}{e^{2\kappa_t}+1}, \quad d\kappa_t = 0.03\left(0.64-\kappa_t\right)dt + 0.118\kappa_t dW_t^{\kappa},$$

$$W_t^{\kappa} = \sqrt{0.96}W_t^{\diamond} - 0.2\sum_{i=1}^{p} W_{i,t}^{X}/\sqrt{p},$$

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

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

  for $1 \le i \ne j \le p$.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

- 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 $1 \le i \ne j \le p$, where $\mathcal{B}$ is a collection of row and column indices (i,j) located within randomly generated diagonal blocks[1].
- Exponentially decaying structure for $\mathbf{\Sigma}_t$ and $\mathbf{\Omega}(t)$: The off-diagonal elements are defined by

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

It is clear that the (exact) sparsity condition is not satisfied when the off-diagonal elements of $\mathbf{\Sigma}_t$ and $\mathbf{\Omega}(t)$ are exponentially decaying as in (5.2). The number of assets p is set as $p = 200$ and 500 and the replication number is $R = 200$.

5.1.2. *Volatility Matrix Estimation.* In the simulation studies, we consider the following volatility matrix estimates.

- Noise-free spot volatility matrix estimate $\widehat{\mathbf{\Sigma}}_t$: 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 $i \ne j$: 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 regularization technique.
- Noise-contaminated spot volatility matrix estimate $\widetilde{\mathbf{\Sigma}}_t$: 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 $i \ne j$ and also compute the naive estimate without applying the shrinkage.
- Time-varying noise volatility matrix estimate $\widehat{\mathbf{\Omega}}(t)$: We combine the kernel smoothing with four shrinkage techniques in the estimation as in Section 3.2 and also the naive estimate without shrinkage.

The choice of tuning parameter in shrinkage is similar to that in Dai et al. (2019). For example, in the noise-free spot volatility estimate, we set the tuning parameter as $\rho_{ij}(t) = \rho(t)(\widehat{\Sigma}_{ii,t}\widehat{\Sigma}_{jj,t})^{1/2}$ 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

[1] As in Dai et al. (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. 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.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

is positive-definite. To evaluate the estimation performance of $\widehat{\mathbf{\Sigma}}_t$, 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:

$$\text{MFL} = \frac{1}{200}\sum_{m=1}^{200}\left(\frac{1}{21}\sum_{j=1}^{21}\left\|\widehat{\mathbf{\Sigma}}_{t_j}^{(m)} - \mathbf{\Sigma}_{t_j}^{(m)}\right\|_F\right),$$

$$\text{MSL} = \frac{1}{200}\sum_{m=1}^{200}\left(\frac{1}{21}\sum_{j=1}^{21}\left\|\widehat{\mathbf{\Sigma}}_{t_j}^{(m)} - \mathbf{\Sigma}_{t_j}^{(m)}\right\|\right),$$

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

5.1.3. *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 on the right. The first two rows of each panel contain the MFL and MSL results for the spot volatility matrix estimation, whereas the third row contains the results for the time-varying noise volatility matrix estimation.

For the noise-free estimate $\widehat{\mathbf{\Sigma}}_t$, when the volatility matrix structure is banding, the performance of the four shrinkage estimators is 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 show that the shrinkage technique is an effective tool in estimating the sparse volatility matrix. Similar results are obtained for the noise-contaminated estimate $\widetilde{\mathbf{\Sigma}}_t$. 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. We next turn the attention to the time-varying noise volatility matrix estimate $\widehat{\mathbf{\Omega}}(t)$. 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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

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

		"Banding"										
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\mathbf{\Sigma}}_t$	MFL	14.396	11.407	5.490	4.038	4.830	MSL	3.963	1.799	1.073	0.867	0.987
$\widetilde{\mathbf{\Sigma}}_t$	MFL	18.497	12.899	12.196	12.064	12.177	MSL	4.796	2.347	2.260	2.255	2.262
$\widehat{\mathbf{\Omega}}(t)$	MFL	11.714	4.226	4.740	3.237	3.960	MSL	3.281	0.682	1.039	0.571	0.753
		"Block-diagonal"										
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\mathbf{\Sigma}}_t$	MFL	14.398	11.277	5.818	4.786	5.424	MSL	4.000	2.293	1.310	1.233	1.386
$\widetilde{\mathbf{\Sigma}}_t$	MFL	18.475	12.811	12.192	12.059	12.158	MSL	4.915	2.777	2.663	2.669	2.662
$\widehat{\mathbf{\Omega}}(t)$	MFL	11.713	4.076	4.875	3.240	3.964	MSL	3.274	0.741	1.098	0.606	0.816
		"Exponentially decaying"										
		Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
$\widehat{\mathbf{\Sigma}}_t$	MFL	14.402	12.033	6.091	5.287	5.976	MSL	4.078	2.456	1.410	1.348	1.510
$\widetilde{\mathbf{\Sigma}}_t$	MFL	18.738	13.464	12.748	12.655	12.739	MSL	4.977	2.934	2.810	2.819	2.815
$\widehat{\mathbf{\Omega}}(t)$	MFL	11.715	4.330	4.860	3.355	4.077	MSL	3.297	0.774	1.085	0.626	0.833

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

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. Overall, the results are very similar to those in Table 1, so we omit the detailed discussion and comparison to save space.

5.2. Simulation for Factor-Based Spot Volatility Matrix Estimation

5.2.1. *Simulation Setup.* We generate $\mathbf{Y}_t$ via (4.1), where the p-dimensional idiosyncratic returns follow the dynamics of $d\mathbf{X}_t$ defined in (5.1). In this simulation, we only consider $p = 500$. As in Aït-Sahalia et al. (2020), we adopt a three-factor model, where the factors $\mathbf{F}_t = \left(F_{1,t}, F_{2,t}, F_{3,t}\right)^\intercal$ are generated by

$$\begin{pmatrix} dF_{1,t} \\ dF_{2,t} \\ dF_{3,t} \end{pmatrix} = \begin{pmatrix} \mu_1^F \\ \mu_2^F \\ \mu_3^F \end{pmatrix} dt + \begin{pmatrix} \sigma_{1,t} & 0 & 0 \\ 0 & \sigma_{2,t} & 0 \\ 0 & 0 & \sigma_{3,t} \end{pmatrix} \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \begin{pmatrix} dW_{1,t}^F \\ dW_{2,t}^F \\ dW_{3,t}^F \end{pmatrix}.$$

The factor volatilities are driven by

$$d\sigma_{k,t}^2 = \tilde{\kappa}_k \left(\tilde{\alpha}_k - \sigma_{k,t}^2\right) dt + \tilde{\nu}_k \sigma_{k,t} d\widetilde{W}_{k,t}, \quad k = 1, 2, 3,$$

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

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

		"Banding"										
		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
$\widehat{\mathbf{\Omega}}(t)$	MFL	18.269	4.045	4.826	5.532	4.547	MSL	3.307	0.461	0.540	0.675	0.519
		"Block-diagonal"										
		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
$\widehat{\mathbf{\Omega}}(t)$	MFL	18.271	4.208	4.935	5.686	4.684	MSL	3.312	0.522	0.603	0.751	0.572
		"Exponentially decaying"										
		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
$\widehat{\mathbf{\Omega}}(t)$	MFL	18.275	4.335	5.001	5.763	4.745	MSL	3.322	0.533	0.610	0.757	0.578

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

where $\mathsf{E}[dW_{k,t}^F d\widetilde{W}_{k,t}] = \rho_k dt$, allowing for potential leverage effects in the factor dynamics. Both $W_{k,t}^F$ and $\widetilde{W}_{k,t}$ are standard univariate Brownian motions. We set $(\tilde{\kappa}_1, \tilde{\kappa}_2, \tilde{\kappa}_3) = (3, 4, 5)$, $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3) = (0.09, 0.04, 0.06)$, $(\tilde{\nu}_1, \tilde{\nu}_2, \tilde{\nu}_3) = (0.3, 0.4, 0.3)$, $\left(\mu_1^F, \mu_2^F, \mu_3^F\right) = (0.05, 0.03, 0.02)$, $(\rho_1, \rho_2, \rho_3) = (-0.6, -0.4, -0.25)$, and $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.05, 0.10, 0.15)$, and consider the following three cases for generating the time-varying beta processes $\beta_i(t) = \left[\beta_{i,1}(t), \beta_{i,2}(t), \beta_{i,3}(t)\right]^\intercal$, $i = 1, \ldots, p$.

- Constant betas. The factor loadings are constants over time, i.e., $\beta_{i,l}(t) = \beta_{i,l}$, $i = 1, \ldots, p$, and $l = 1, 2, 3$. For each i, we set $\beta_{i,1} \sim \mathsf{U}(0.25, 2.25)$ and $\beta_{i,2}, \beta_{i,3} \sim \mathsf{U}(-0.5, 0.5)$.
- Deterministic time-varying betas. Consider the following deterministic function:

$$\beta_{i,l}(t) = \frac{1}{2}\left[\cos\left(\pi(t - \omega_{i,l})/T\right) + 1\right] \times \left(\overline{\beta}_{i,l} - \underline{\beta}_{i,l}\right) + \underline{\beta}_{i,l}, \quad i = 1, \ldots, p, \quad l = 1, 2, 3,$$

  where $\omega_{i,1}, \omega_{i,2}, \omega_{i,3} \sim \mathsf{U}(0, 2T)$, $(\underline{\beta}_{i,1}, \overline{\beta}_{i,1})$ is a pair of random numbers from $\mathsf{U}(0.25, 2.25)$ whereas $(\underline{\beta}_{i,2}, \overline{\beta}_{i,2})$ and $(\underline{\beta}_{i,3}, \overline{\beta}_{i,3})$ are pairs of random numbers from $\mathsf{U}(-0.5, 0.5)$.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

- Stochastic time-varying betas. As in Aït-Sahalia et al. (2020), we consider the following diffusion process:

$$d\beta_{i,l}(t) = \kappa^{\beta}_{i,l}\left(\alpha^{\beta}_{i,l} - \beta_{i,l}(t)\right)dt + \upsilon^{\beta}_{i,l}dW^{\beta}_{i,l,t}, \quad i = 1,\ldots,p, \quad l = 1,2,3,$$

where $W^{\beta}_{i,l,t}$ are standard Brownian motions independently over i and l, $\kappa^{\beta}_{i,1}, \kappa^{\beta}_{i,2}, \kappa^{\beta}_{i,3} \sim \mathsf{U}(1,3)$, $\alpha^{\beta}_{i,1} \sim \mathsf{U}(0.25, 2.25)$, $\alpha^{\beta}_{i,2}, \alpha^{\beta}_{i,3} \sim \mathsf{U}(-0.5, 0.5)$, and $\upsilon^{\beta}_{i,1}, \upsilon^{\beta}_{i,2}, \upsilon^{\beta}_{i,3} \sim \mathsf{U}(2,4)$.

We also generate the noise-contaminated version of $\mathbf{Y}_t$ and $\mathbf{F}_t$ as in Section 5.1.1, i.e., $\boldsymbol{\xi}^{*}_{Y,k}$ and $\boldsymbol{\xi}^{*}_{F,k}$ are independent vectors of cross-sectionally independent standard normal random variables, and the noise covariance structure satisfies the banding, block-diagonal, or exponential decaying assumption as for $\boldsymbol{\Omega}(\cdot)$.

5.2.2. *Simulation Results.* For ease of comparison, we use exactly the same bandwidth as in our first experiment. The results for the noise-free and noise-contaminated spot idiosyncratic volatility matrix estimates $\widehat{\boldsymbol{\Sigma}}^{X}_{t}$ and $\widetilde{\boldsymbol{\Sigma}}^{X}_{t}$ measured by MFL and MSL are reported in Table 3, which reveal some desirable observations. First, we note that our estimation results in terms of MFL and MSL are almost identical across different types of dynamics of factor loadings, indicating that the developed estimation procedure is robust in finite samples to different assumptions of the factor loading dynamics as long as they satisfy our smooth restriction (see Assumption 6(iv)). Second, the MFL and MSL values are similar to those reported in Table 2, which were obtained based on the data generating model without common factors. This means that the proposed nonparametric time-varying high-frequency regression can effectively remove common factors, resulting in accurate estimation of the spot idiosyncratic volatility matrix.

Let $\widehat{\boldsymbol{\Sigma}}^{Y}_{t}$ and $\widetilde{\boldsymbol{\Sigma}}^{Y}_{t}$ denote the noise-free and noise-contaminated factor-based spot volatility matrix estimates, respectively. As discussed in Section 4, we measure the accuracy of the spiked volatility matrix estimate by the relative error defined above Theorem 4, i.e., consider the following Mean Relative Loss (MRL):

$$\text{MRL} = \frac{1}{200}\sum_{m=1}^{200}\left(\frac{1}{21}\sum_{j=1}^{21}\left\|\overline{\boldsymbol{\Sigma}}^{Y,(m)}_{t_j} - \boldsymbol{\Sigma}^{Y,(m)}_{t_j}\right\|_{\boldsymbol{\Sigma}^{Y,(m)}_{t_j}}\right),$$

where $\overline{\boldsymbol{\Sigma}}^{Y,(m)}_{t}$ denotes $\widehat{\boldsymbol{\Sigma}}^{Y}_{t}$ or $\widetilde{\boldsymbol{\Sigma}}^{Y}_{t}$ in the mth replication. The relevant results are reported in Table 4. We can see that the performance of the shrinkage estimates is substantially better than that of the naive estimate. Unsurprisingly, due to the presence of microstructure noise, the MRL results of $\widetilde{\boldsymbol{\Sigma}}^{Y}_{t}$ are much higher than those of $\widehat{\boldsymbol{\Sigma}}^{Y}_{t}$. As in Table 3, our proposed estimation is robust to different factor loading dynamics.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

TABLE 3. Estimation results for the spot idiosyncratic volatility matrices.

			"Banding"										
β Dynamics			Frobenius norm						Spectral norm				
			Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
Constant	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9037	4.2461	5.2485	4.9880	3.9910	MSL	3.8887	0.6359	0.7291	0.7154	0.5720
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.6646	19.3752	18.3036	17.7552	18.1388	MSL	11.1910	2.3576	2.2653	2.2160	2.2554
Deterministic	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9127	4.1916	5.2503	4.9898	3.9842	MSL	3.8901	0.6313	0.7288	0.7144	0.5712
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.5672	19.3633	18.2947	17.7267	18.1284	MSL	10.9662	2.3571	2.2636	2.2128	2.2536
Stochastic	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9099	4.2123	5.2498	4.9893	3.9872	MSL	3.8896	0.6331	0.7289	0.7149	0.5717
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.7323	19.3896	18.3164	17.7839	18.1538	MSL	11.3262	2.3603	2.2708	2.2203	2.2602
			"Block-diagonal"										
β Dynamics			Frobenius norm						Spectral norm				
			Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
Constant	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9047	5.6802	6.4718	5.9421	5.4710	MSL	3.9741	0.8722	1.1481	0.9106	0.9014
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.7266	19.8195	18.8114	18.3162	18.6638	MSL	10.9751	2.8701	2.7656	2.7097	2.7551
Deterministic	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9137	5.6821	6.4738	5.9436	5.4729	MSL	3.9754	0.8718	1.1479	0.9103	0.9012
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.6284	19.8161	18.8043	18.2953	18.6547	MSL	10.7452	2.8706	2.7663	2.7092	2.7559
Stochastic	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9108	5.6811	6.4732	5.9433	5.4722	MSL	3.9751	0.8721	1.1480	0.9104	0.9013
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.7955	19.8314	18.8237	18.3434	18.6767	MSL	11.1149	2.8719	2.7691	2.7142	2.7584

(Continued)

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

TABLE 3. Continued

β Dynamics			"Exponentially decaying"										
			Frobenius norm						Spectral norm				
			Naive	Hard	Soft	AL	SCAD		Naive	Hard	Soft	AL	SCAD
Constant	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9057	6.0626	6.7715	6.1573	5.7617	MSL	4.0142	0.9106	1.1898	0.9453	0.9388
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.8728	20.3802	19.2709	18.7715	19.1262	MSL	10.8858	2.9381	2.8260	2.7707	2.8154
Deterministic	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9147	6.0709	6.7737	6.1589	5.7637	MSL	4.0156	0.9112	1.1896	0.9450	0.9387
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.7746	20.3564	19.2632	18.7460	19.1173	MSL	10.6538	2.9354	2.8247	2.7673	2.8140
Stochastic	$\widehat{\mathbf{\Sigma}}_t^X$	MFL	21.9118	6.0636	6.7730	6.1585	5.7630	MSL	4.0151	0.9106	1.1897	0.9451	0.9388
	$\widetilde{\mathbf{\Sigma}}_t^X$	MFL	30.9430	20.3820	19.2839	18.8017	19.1405	MSL	11.0295	2.9381	2.8291	2.7745	2.8173

TABLE 4. Mean relative loss for the factor-based spot volatility matrix estimation.

		"Banding"				
β Dynamics		Naive	Hard	Soft	AL	SCAD
Constant	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1192	0.5417	0.7802	0.7762	0.4391
	$\widetilde{\mathbf{\Sigma}}_t^Y$	2.2280	1.7243	1.4939	1.4478	1.4654
Deterministic	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1207	0.5257	0.7823	0.7775	0.4371
	$\widetilde{\mathbf{\Sigma}}_t^Y$	2.2287	1.7182	1.4882	1.4385	1.4586
Stochastic	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1208	0.5389	0.7829	0.7780	0.4406
	$\widetilde{\mathbf{\Sigma}}_t^Y$	2.2273	1.7279	1.4986	1.4544	1.4719
		"Block-diagonal"				
β Dynamics		Naive	Hard	Soft	AL	SCAD
Constant	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1192	0.3842	0.3650	0.3962	0.3249
	$\widetilde{\mathbf{\Sigma}}_t^Y$	1.7146	0.8421	0.7938	0.7176	0.7514
Deterministic	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1201	0.3840	0.3651	0.3958	0.3241
	$\widetilde{\mathbf{\Sigma}}_t^Y$	1.7152	0.8410	0.7911	0.7155	0.7486
Stochastic	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1202	0.3868	0.3678	0.3983	0.3272
	$\widetilde{\mathbf{\Sigma}}_t^Y$	1.7146	0.8435	0.7949	0.7188	0.7528
		"Exponentially decaying"				
β Dynamics		Naive	Hard	Soft	AL	SCAD
Constant	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1192	0.4086	0.3726	0.4055	0.3347
	$\widetilde{\mathbf{\Sigma}}_t^Y$	1.7338	0.8636	0.8047	0.7272	0.7619
Deterministic	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1201	0.4079	0.3727	0.4051	0.3339
	$\widetilde{\mathbf{\Sigma}}_t^Y$	1.7344	0.8614	0.8016	0.7249	0.7589
Stochastic	$\widehat{\mathbf{\Sigma}}_t^Y$	1.1203	0.4111	0.3754	0.4075	0.3370
	$\widetilde{\mathbf{\Sigma}}_t^Y$	1.7338	0.8645	0.8058	0.7283	0.7631

6. EMPIRICAL STUDY

We apply the proposed methods to the intraday returns of the S&P 500 component stocks to demonstrate the effectiveness of our nonparametric spot volatility matrix estimation in revealing time-varying patterns. We consider the 5-minute returns of the S&P 500 stocks collected in September 2008. On September 15, Lehman Brothers filed for bankruptcy, causing shockwaves throughout the global financial system. Hence, it is interesting to examine how the spot volatility structure of the returns evolved during this one-month period. In addition, to demonstrate the effectiveness of our model with observed risk factors in explaining the systemic component of the dependence structure, we also collect the 5-minute returns of 12 factors. The first three factors are constructed in Aït-Sahalia et al. (2020) as

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

our proxy for the market (MKT), small-minus-big market capitalization (SMB), and high-minus-low price-earning ratio (HML). The other nine factors are the widely available sector SDPR ETFs, which are intended to tract the following nine largest S&P sectors: Energy (XLE), Materials (XLB), Industrials (XLI), Consumer Discretionary (XLY), Consumer Staples (XLP), Health Care (XLV), Financial (XLF), Information Technology (XLK), and Utilities (XLU). We sort our stocks according to their GICS (Global Industry Classification Standard) codes, so that they are grouped by sectors in the above order. Consequently, the correlation (sub)matrix for stocks within each sector corresponds to a block on the diagonal of the full correlation matrix (e.g., Fan et al., 2016).

We only use stocks that are included in the S&P 500 index and whose GICS codes are unchanged in September 2008. We also exclude stocks that do not belong to any of the above nine sectors. This leaves us with a total of $p = 482$ stocks. All the returns are synchronized via the previous-tick subsampling technique (Zhang, 2011), and overnight returns are removed because of potential dividends and stock splits. Consequently, we have $1,638$ time-series observations for each of the 482 stocks. For the 5-minute returns, we may assume that the potential impact of microstructure noise is negligible.[2] The smoothing parameter in our kernel estimation is chosen as $h = 2/252$ (equivalent to 2 trading days)[3].

We start with estimating the spot volatility matrices of the total returns (i.e., the observed returns) without incorporating the observed factors or applying any shrinkage. To visualize the potential time variation of the estimated spot matrices, as in Bibinger et al. (2019), we plot the time series of deciles of the distribution of the estimated spot variances and the pairwise correlations.[4] The patterns of the spot variances and correlations in Figure 1a,b reveal some clear evidence of time variation in our sampling period. We note that the distributions of the variances are relatively narrow and stay low on the first few days of the month. However, close to Lehman Brothers' announcement on the 15th, they start to rise and get wider quite rapidly and reach the peak around the 17th and the 18th. The spot variances at the peak are much higher than those on the earlier days of the month. The distributions return to the earlier level in the following week. In contrast, the distributions of pairwise spot correlations also start to shift up around the same time, but quickly reach the peak on the 16th (only one day after the bankruptcy news), and then dip to a relatively low point around the 19th before returning to the earlier level. Such time-varying features in the dynamic covariance structure

[2]It is possible to estimate the spot volatility matrix of the microstructure noise vector. However, since the focus of this paper is on the spot volatility structure of the returns and the microstructure noise for 5-minute returns is negligible, we choose not to investigate in this section but to leave this exercise for future research.

[3]We tested three bandwidth choices, namely, 1 day, 2 days, and 3 days, and found that $h = 1/252$ ($h = 3/252$) produced clearly under-smoothed (over-smoothed) time series of estimated deciles of the cross-sectional distribution of the variances and pairwise correlations of our returns, whereas $h = 2/252$ seems to be the most reasonable one. Our qualitative conclusion is unaffected by the choices of h within the range of 1–3 days. A smaller bandwidth (e.g., 2 hours), applied to data with a higher frequency (e.g., 1 minute or shorter), may be capable of revealing some intraday patterns of the spot volatility dynamics (e.g., Andersen et al., 2024).

[4]The nine decile levels we use in this study are the 10th, 20th,…, and 90th percentiles.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

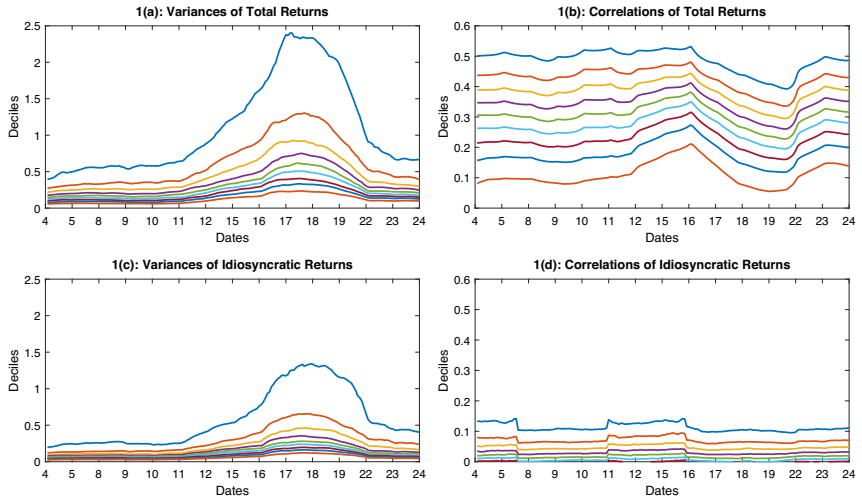


FIGURE 1. Deciles of spot variances and pairwise correlations in September 2008.

are quite interesting and sensible, reflecting the impact of market news. Hence, our proposed spot volatility matrix estimation methodology provides a useful tool for revealing such dynamics.

To examine whether it is appropriate to directly apply shrinkage techniques to the spot volatility matrices of the total returns, following Fan et al. (2016), we plot in Figure 2a,b their sparsity patterns on the 16th and the 19th of September. The deep blue dots correspond to the locations of pairwise correlations that are at least 0.15, whereas the white dots correspond to those smaller than 0.15. Note that the covariance structure of the total returns is very dense on these two days. Therefore, it is not appropriate to directly apply the shrinkage technique as in Sections 2 and 3. Meanwhile, although both are quite dense, we can still clearly see their differences. Consistent with our observation from the decile plots of the correlations, we can see that the plot for the 16th is almost completely covered by blue dots, but the plot for the 19th has significantly more areas covered in white.

We next incorporate the 12 observed factors in the large spot volatility matrix estimation as suggested in Section 4. In particular, we are interested in estimating the spot idiosyncratic volatility matrix, which is expected to satisfy the sparsity restriction. To save space, we choose to only report results using the SCAD shrinkage due to its satisfactory performance in our simulation. In Figure 1c,d, we plot the deciles of the estimated spot idiosyncratic variances and correlations over trading days. In Figure 1c, we observe a significant upward shift of the distribution of the spot variances of the idiosyncratic returns around the time of Lehman Brothers' bankruptcy, indicating that the observed factors may not fully capture the time variation of the spot variances. In contrast, the deciles of the spot correlations

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

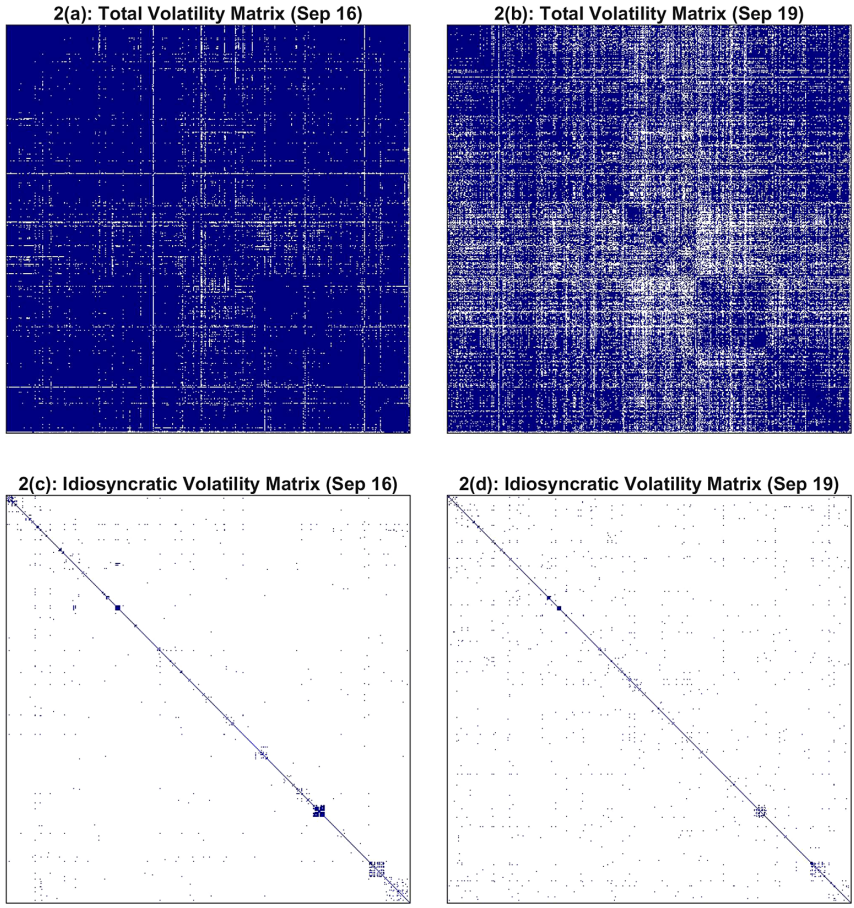


FIGURE 2. Sparsity patterns of the total and idiosyncratic volatility matrix estimates on September 16 and 19, 2008.

in Figure 1d seem to be quite flat throughout the entire month, suggesting that the systematic factors may explain the time variation in the distribution of the pairwise correlations better than that of the variances.

We finally plot the sparsity patterns of the two estimated spot idiosyncratic volatility matrices on September 16 and 19 in Figure 2c,d, respectively. Unlike Figure 2a,b, we note that the estimated spot idiosyncratic volatility matrices are highly sparse on both days. This is consistent with our observation from Figure 1d, confirming that the observed factors can effectively account for the time variation in the spot covariance structure of the returns. Meanwhile, we also note that the two idiosyncratic volatility matrices are clearly not diagonal and still carry some visible time variation. Lastly, it is worth mentioning that the estimated spot idiosyncratic

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

volatility matrices do not exhibit significant correlations within the blocks along the diagonal lines, except for some very limited actions in the lower right corner of the two matrices, which corresponds to the XLU sector.

7. CONCLUSION

We develop the nonparametric kernel-weighted smoothing combined with the generalized shrinkage to estimate the high-dimensional spot volatility and time-varying noise covariance matrices under the uniform sparsity assumption. A localized pre-averaging method is proposed to accommodate the market microstructure noise. We further estimate the spot volatility matrix with observed common risk factors, relaxing the sparse structural restriction. The uniform consistency properties with convergence rates are derived for the proposed matrix estimates. The simulation results show that the estimation methods work well in finite samples for both the noise-free and noise-contaminated data and the empirical study demonstrates the effectiveness of the developed model and method on the 5-minute returns of S&P 500 stocks.

The methodology and theory developed in Sections 2–4 rule out jumps in prices and volatility, which are not uncommon in practice when the high-frequency financial data are collected for a large number of assets. There have been extensive studies on testing and estimation of jumps in high-frequency data (e.g., Lee and Mykland, 2008; Jacod and Todorov, 2010; Aït-Sahalia, Jacod, and Li, 2012; Bibinger, Jirak, and Vetter, 2017; Li, Todorov, and Tauchen, 2017, 2019). We next briefly discuss the modification of the proposed high-dimensional spot volatility matrix estimation to accommodate price jumps. For notational brevity, we only consider the noise-free high-frequency data as in Section 2, and the amendment is similar for the estimators in Sections 3 and 4. We replace the stochastic differential equation (2.1) by

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

where $\mathbf{J}_t$ is a p-dimensional pure jump process, i.e., $\mathbf{J}_t = \sum_{s \le t} \Delta \mathbf{J}_s$ where $\Delta \mathbf{J}_s$ denotes jumps of the process. Assume that the price jumps are of finite activity for convenience. As recommended by Mancini (2009) and Jacod and Protter (2012), we adopt a standard truncation to eliminate price jumps in estimating $\Sigma_{ij,t}$. Define

$$\mathcal{T}_{ij,k} = I\left(\left|\Delta X_{i,k}\right| \le \Delta^{\varphi},\ \left|\Delta X_{j,k}\right| \le \Delta^{\varphi}\right),$$

where $0 < \varphi < 1/2$. Then, we may estimate $\Sigma_{ij,t}$ by

$$\widehat{\Sigma}^{\mathcal{T}}_{ij,t} = \sum_{k=1}^{n} K_h^*(t_k - t)\Delta X_{i,k}\Delta X_{j,k}\mathcal{T}_{ij,k}.$$

A generalized shrinkage is then applied to $\widehat{\Sigma}^{\mathcal{T}}_{ij,t}$, $i \neq j$, as in (2.5). We conjecture that Theorem 1 continues to hold when there exist price jumps. Furthermore, when there are possible jumps in volatility, as suggested by Bibinger et al. (2017), we

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

may use the one-sided kernel smoothing to estimate the left and right limits of the spot volatility and subsequently construct a suitable test statistic. It is nontrivial to extend the methodology and theory in Bibinger et al. (2017) to the high-dimensional setting. We will explore it in our future study.

APPENDIX

A. Proofs of the Main Results

In this appendix, we give the proofs of the main theorems. We start with four propositions whose proofs are available in Appendix B of the Supplementary Material.

PROPOSITION A.1. *Suppose that Assumptions 1 and 2(i) and (ii) are satisfied. Then, we have*

$$\max_{1\le i,j\le p}\sup_{0\le t\le T}\left|\widehat{\Sigma}_{ij,t}-\Sigma_{ij,t}\right| = O_P\left(\zeta_{\Delta,p}\right), \tag{A.1}$$

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

PROPOSITION A.2. *Suppose that Assumptions 1, 2(i), 3, and 4(i) and (ii) are satisfied, and Assumption 2(ii) holds with* Δ^{-1} *replaced by N.*

$$\max_{1\le i,j\le p}\sup_{0\le t\le T}\left|\widetilde{\Sigma}_{ij,t}-\Sigma_{ij,t}\right| = O_P\left(\zeta^{*}_{N,p}+\nu_{\Delta,p,N}\right), \tag{A.2}$$

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. Then, we have*

$$\max_{1\le i,j\le p}\sup_{0\le t\le T}\left|\widehat{\Omega}_{ij}(t)-\Omega_{ij}(t)\right| = O_P\left(\delta_{\Delta,p}\right), \tag{A.3}$$

where $\delta_{\Delta,p} = h_1^{\gamma_1} + \left[\frac{\Delta\log(p\vee\Delta^{-1})}{h_1}\right]^{1/2}$.

PROPOSITION A.4. *Suppose that Assumptions 2(i) and (ii) and 6 are satisfied. Then, we have*

$$\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}^{Y}_{t}-\mathbf{\Sigma}^{Y}_{t}\right\|_{\max} = O_P\left(\zeta_{\Delta,p}\right),$$
$$\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}^{F}_{t}-\mathbf{\Sigma}^{F}_{t}\right\|_{\max} = O_P\left(\zeta_{\Delta,p}\right),$$
$$\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}^{YF}_{t}-\mathbf{\Sigma}^{YF}_{t}\right\|_{\max} = O_P\left(\zeta_{\Delta,p}\right),$$

where $\zeta_{\Delta,p}$ *is defined as in Proposition A.1.*

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

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

$$\begin{aligned}
&\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}_t^s-\mathbf{\Sigma}_t\right\| \le \sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1}^{p}\left|\widehat{\Sigma}_{ij,t}^s-\Sigma_{ij,t}\right| \\
&= \sup_{0\le t\le T}\max_{1\le i\le p}\left|\widehat{\Sigma}_{ii,t}-\Sigma_{ii,t}\right| + \sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1,\neq i}^{p}\left|s_{\rho_1(t)}\left(\widehat{\Sigma}_{ij,t}\right)I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)-\Sigma_{ij,t}\right| \\
&= \sup_{0\le t\le T}\max_{1\le i\le p}\left|\widehat{\Sigma}_{ii,t}-\Sigma_{ii,t}\right| + \sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1,\neq i}^{p}\left|s_{\rho_1(t)}\left(\widehat{\Sigma}_{ij,t}\right)I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)-\right. \\
&\quad \left.\Sigma_{ij,t}I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)-\Sigma_{ij,t}I\left(\left|\widehat{\Sigma}_{ij,t}\right|\le\rho_1(t)\right)\right| \\
&\le \sup_{0\le t\le T}\max_{1\le i\le p}\left|\widehat{\Sigma}_{ii,t}-\Sigma_{ii,t}\right| + \sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1,\neq i}^{p}\left|s_{\rho_1(t)}\left(\widehat{\Sigma}_{ij,t}\right)-\widehat{\Sigma}_{ij,t}\right|I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)+ \\
&\quad \sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1,\neq i}^{p}\left|\widehat{\Sigma}_{ij,t}-\Sigma_{ij,t}\right|I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)+\sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1,\neq i}^{p}\left|\Sigma_{ij,t}\right|I\left(\left|\widehat{\Sigma}_{ij,t}\right|\le\rho_1(t)\right) \\
&=: \Pi_1+\Pi_2+\Pi_3+\Pi_4. \qquad \textbf{(A.4)}
\end{aligned}$$

Define the event

$$\mathcal{G}(M)=\left\{\max_{1\le i,j\le p}\sup_{0\le t\le T}\left|\widehat{\Sigma}_{ij,t}-\Sigma_{ij,t}\right|\le M\zeta_{\Delta,p}\right\},$$

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

$$\mathsf{P}\left(\mathcal{G}(M_\epsilon)\right)\ge 1-\epsilon. \qquad \textbf{(A.5)}$$

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

$$\Pi_2\le \sup_{0\le t\le T}\rho_1(t)\left[\max_{1\le i\le p}\sum_{j=1}^{p}I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)\right]$$

and

$$\Pi_3\le M_\epsilon\zeta_{\Delta,p}\left[\sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1}^{p}I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\rho_1(t)\right)\right]$$

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

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

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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

$$\begin{aligned}\Pi_2+\Pi_3 &\le \zeta_{\Delta,p}(\overline{C}_M+M_\epsilon)\left[\sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1}^{p} I\left(\left|\widehat{\Sigma}_{ij,t}\right|>\underline{C}_M\zeta_{\Delta,p}\right)\right]\\ &\le \zeta_{\Delta,p}(\overline{C}_M+M_\epsilon)\left[\sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1}^{p} I\left(\left|\Sigma_{ij,t}\right|>M_\epsilon\zeta_{\Delta,p}\right)\right]\\ &= O_P\left(\zeta_{\Delta,p}\right)\left[\sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1}^{p}\frac{|\Sigma_{ij,t}|^q}{\left(M_\epsilon\zeta_{\Delta,p}\right)^q}\right]\\ &= O_P\left(\Lambda\varpi(p)\zeta_{\Delta,p}^{1-q}\right)=O_P\left(\varpi(p)\zeta_{\Delta,p}^{1-q}\right) \qquad \textbf{(A.6)}\end{aligned}$$

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

$$\begin{aligned}\Pi_4 &\le \sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_{ij,t}|I\left(|\Sigma_{ij,t}|\le\left(\overline{C}_M+M_\epsilon\right)\zeta_{\Delta,p}\right)\\ &\le \left(\overline{C}_M+M_\epsilon\right)^{1-q}\zeta_{\Delta,p}^{1-q}\sup_{0\le t\le T}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_{ij,t}|^q\\ &= O_P\left(\Lambda\varpi(p)\zeta_{\Delta,p}^{1-q}\right)=O_P\left(\varpi(p)\zeta_{\Delta,p}^{1-q}\right). \qquad \textbf{(A.7)}\end{aligned}$$

By Proposition A.1, we readily have that

$$\Pi_1 = O_P\left(\zeta_{\Delta,p}\right)=O_P\left(\varpi(p)\zeta_{\Delta,p}^{1-q}\right). \qquad \textbf{(A.8)}$$

By (A.6)–(A.8), 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 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 space. □

Proof of Theorem 4. By Proposition A.4 and the definition of $\widehat{\Sigma}_{ij,t}^X$ in (4.8), we may show that

$$\max_{1\le i,j\le p}\sup_{0\le t\le T}\left|\widehat{\Sigma}_{ij,t}^X-\Sigma_{ij,t}^X\right|=O_P\left(\zeta_{\Delta,p}\right). \qquad \textbf{(A.9)}$$

With (A.9), following the proof of Theorem 1, we complete the proof of (4.12).

We next turn to the proof of (4.13). Note that

$$\sup_{0\le t\le T}\left\|\widehat{\boldsymbol{\Sigma}}_t^{Y,s}-\boldsymbol{\Sigma}_t^Y\right\|_{\boldsymbol{\Sigma}_t^Y}^2\le 2\sup_{0\le t\le T}\left[\left\|\widehat{\boldsymbol{\Sigma}}_t^{X,s}-\boldsymbol{\Sigma}_t^X\right\|_{\boldsymbol{\Sigma}_t^Y}^2+\left\|\widehat{\boldsymbol{\beta}}(t)\widehat{\boldsymbol{\Sigma}}_t^F\widehat{\boldsymbol{\beta}}(t)^\intercal-\boldsymbol{\beta}(t)\boldsymbol{\Sigma}_t^F\boldsymbol{\beta}(t)^\intercal\right\|_{\boldsymbol{\Sigma}_t^Y}^2\right].$$

For any $p\times p$ matrix $\boldsymbol{\Sigma}$, since all the eigenvalues of $\boldsymbol{\Sigma}_t^Y$ are strictly larger than a positive constant,

$$\|\boldsymbol{\Sigma}\|_{\boldsymbol{\Sigma}_t^Y}^2=\frac{1}{p}\left\|\left(\boldsymbol{\Sigma}_t^Y\right)^{-1/2}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}_t^Y\right)^{-1/2}\right\|_F^2\le\frac{C}{p}\|\boldsymbol{\Sigma}\|_F^2, \qquad \textbf{(A.10)}$$

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

where $C > 0$ is a generic constant whose value may change from line to line. By (4.12) and (A.10), we prove

$$\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}_t^{X,s}-\mathbf{\Sigma}_t^X\right\|_{\mathbf{\Sigma}_t^Y}^2\le\frac{C}{p}\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}_t^{X,s}-\mathbf{\Sigma}_t^X\right\|_F^2\le C\sup_{0\le t\le T}\left\|\widehat{\mathbf{\Sigma}}_t^{X,s}-\mathbf{\Sigma}_t^X\right\|^2=O_P\left([\varpi(p)\zeta_{\Delta,p}^{1-q}]^2\right). \tag{A.11}$$

By the definition of $\widehat{\boldsymbol{\beta}}(t)$ in (4.7) and Proposition A.4, we readily have that

$$\max_{1\le i\le p}\sup_{0\le t\le T}\left\|\widehat{\beta}_i(t)-\beta_i(t)\right\|_2=O_P\left(\zeta_{\Delta,p}\right). \tag{A.12}$$

Write $\mathbf{D}_t^\beta=\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}(t)$ and $\mathbf{D}_t^F=\widehat{\mathbf{\Sigma}}_t^F-\mathbf{\Sigma}_t^F$. Note that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(t)\widehat{\mathbf{\Sigma}}_t^F\widehat{\boldsymbol{\beta}}(t)^\intercal-\boldsymbol{\beta}(t)\mathbf{\Sigma}_t^F\boldsymbol{\beta}(t)^\intercal&=\mathbf{D}_t^\beta\mathbf{D}_t^F\mathbf{D}_t^{\beta\intercal}+\mathbf{D}_t^\beta\mathbf{\Sigma}_t^F\mathbf{D}_t^{\beta\intercal}+\mathbf{D}_t^\beta\mathbf{D}_t^F\boldsymbol{\beta}(t)^\intercal+\mathbf{D}_t^\beta\mathbf{\Sigma}_t^F\boldsymbol{\beta}(t)^\intercal\\&\quad+\boldsymbol{\beta}(t)\mathbf{D}_t^F\mathbf{D}_t^{\beta\intercal}+\boldsymbol{\beta}(t)\mathbf{\Sigma}_t^F\mathbf{D}_t^{\beta\intercal}+\boldsymbol{\beta}(t)\mathbf{D}_t^F\boldsymbol{\beta}(t)^\intercal.\end{aligned}$$

By (A.10), (A.12), and Proposition A.4, we have

$$\begin{aligned}\sup_{0\le t\le T}\left\|\mathbf{D}_t^\beta\mathbf{D}_t^F\mathbf{D}_t^{\beta\intercal}\right\|_{\mathbf{\Sigma}_t^Y}^2&\le C\sup_{0\le t\le T}\frac{1}{p}\left\|\mathbf{D}_t^\beta\mathbf{D}_t^F\mathbf{D}_t^{\beta\intercal}\right\|_F^2\\&\le\frac{C}{p}\sup_{0\le t\le T}\left\|\mathbf{D}_t^\beta\right\|_F^4\sup_{0\le t\le T}\left\|\mathbf{D}_t^F\right\|^2\\&=O_P\left(p\zeta_{\Delta,p}^6\right).\end{aligned} \tag{A.13}$$

Similarly, we can show that

$$\sup_{0\le t\le T}\left\|\mathbf{D}_t^\beta\mathbf{\Sigma}_t^F\mathbf{D}_t^{\beta\intercal}\right\|_{\mathbf{\Sigma}_t^Y}^2\le C\sup_{0\le t\le T}\frac{1}{p}\left\|\mathbf{D}_t^\beta\right\|_F^4=O_P\left(p\zeta_{\Delta,p}^4\right). \tag{A.14}$$

By (4.3), Assumption 6(iv) and the Sherman–Morrison–Woodbury formula, we may show that

$$\sup_{0\le t\le T}\left\|\boldsymbol{\beta}(t)^\intercal\left(\mathbf{\Sigma}_t^Y\right)^{-1}\boldsymbol{\beta}(t)\right\|=O_P(1). \tag{A.15}$$

Using (A.12), (A.15), and Proposition A.4, we have

$$\begin{aligned}\sup_{0\le t\le T}\left\|\mathbf{D}_t^\beta\mathbf{D}_t^F\boldsymbol{\beta}(t)^\intercal\right\|_{\mathbf{\Sigma}_t^Y}^2&=\frac{1}{p}\sup_{0\le t\le T}\operatorname{trace}\left\{\mathbf{D}_t^F\mathbf{D}_t^{\beta\intercal}\left(\mathbf{\Sigma}_t^Y\right)^{-1}\mathbf{D}_t^\beta\mathbf{D}_t^F\boldsymbol{\beta}(t)^\intercal\left(\mathbf{\Sigma}_t^Y\right)^{-1}\boldsymbol{\beta}(t)\right\}\\&\le\frac{C}{p}\sup_{0\le t\le T}\left\|\mathbf{D}_t^\beta\right\|_F^2\sup_{0\le t\le T}\left\|\mathbf{D}_t^F\right\|^2\sup_{0\le t\le T}\left\|\boldsymbol{\beta}(t)^\intercal\left(\mathbf{\Sigma}_t^Y\right)^{-1}\boldsymbol{\beta}(t)\right\|\\&\le\frac{C}{p}\sup_{0\le t\le T}\left\|\mathbf{D}_t^\beta\right\|_F^2\sup_{0\le t\le T}\left\|\mathbf{D}_t^F\right\|^2=O_P\left(\zeta_{\Delta,p}^4\right)\end{aligned} \tag{A.16}$$

and

$$\sup_{0\le t\le T}\left\|\boldsymbol{\beta}(t)\mathbf{D}_t^F\mathbf{D}_t^{\beta\intercal}\right\|_{\mathbf{\Sigma}_t^Y}^2=O_P\left(\zeta_{\Delta,p}^4\right). \tag{A.17}$$

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

Similar to the proof of (A.16), we also have

$$\sup_{0\le t\le T}\left\|\mathbf{D}_t^{\beta}\boldsymbol{\Sigma}_t^F\boldsymbol{\beta}(t)^{\intercal}\right\|_{\boldsymbol{\Sigma}_t^Y}^2 \le \frac{C}{p}\sup_{0\le t\le T}\left\|\mathbf{D}_t^{\beta}\right\|_F^2\sup_{0\le t\le T}\left\|\boldsymbol{\Sigma}_t^F\right\|^2 = O_P\left(\zeta_{\Delta,p}^2\right) \tag{A.18}$$

and

$$\sup_{0\le t\le T}\left\|\boldsymbol{\beta}(t)\boldsymbol{\Sigma}_t^F\mathbf{D}_t^{\beta\intercal}\right\|_{\boldsymbol{\Sigma}_t^Y}^2 = O_P\left(\zeta_{\Delta,p}^2\right). \tag{A.19}$$

By (A.15) and Proposition A.4, we may show that

$$\begin{aligned}\sup_{0\le t\le T}\left\|\boldsymbol{\beta}(t)\mathbf{D}_t^F\boldsymbol{\beta}(t)^{\intercal}\right\|_{\boldsymbol{\Sigma}_t^Y}^2 &= \frac{1}{p}\sup_{0\le t\le T}\operatorname{trace}\left\{\mathbf{D}_t^F\boldsymbol{\beta}(t)^{\intercal}\left(\boldsymbol{\Sigma}_t^Y\right)^{-1}\boldsymbol{\beta}(t)\mathbf{D}_t^F\boldsymbol{\beta}(t)^{\intercal}\left(\boldsymbol{\Sigma}_t^Y\right)^{-1}\boldsymbol{\beta}(t)\right\}\\ &\le \frac{C}{p}\sup_{0\le t\le T}\left\|\mathbf{D}_t^F\right\|^2\sup_{0\le t\le T}\left\|\boldsymbol{\beta}(t)^{\intercal}\left(\boldsymbol{\Sigma}_t^Y\right)^{-1}\boldsymbol{\beta}(t)\right\|^2 = O_P\left(\zeta_{\Delta,p}^2/p\right).\end{aligned} \tag{A.20}$$

With (A.13), (A.14), and (A.16)–(A.20), we have

$$\sup_{0\le t\le T}\left\|\widehat{\boldsymbol{\beta}}(t)\widehat{\boldsymbol{\Sigma}}_t^F\widehat{\boldsymbol{\beta}}(t)^{\intercal}-\boldsymbol{\beta}(t)\boldsymbol{\Sigma}_t^F\boldsymbol{\beta}(t)^{\intercal}\right\|_{\boldsymbol{\Sigma}_t^Y}^2 = O_P\left(p\zeta_{\Delta,p}^4+\zeta_{\Delta,p}^2\right). \tag{A.21}$$

By virtue of (A.11) and (A.21), we complete the proof of (4.13). □

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266466624000264.

REFERENCES

Aït-Sahalia, Y., & Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.

Aït-Sahalia, Y., Jacod, J., & Li, J. (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics*, 168, 207–222.

Aït-Sahalia, Y., Kalnina, I., & Xiu, D. (2020). High-frequency factor models and regressions. *Journal of Econometrics*, 216, 86–105.

Aït-Sahalia, Y., & 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., & Xiu, D. (2019). Principal component analysis of high-frequency data. *Journal of the American Statistical Association*, 114, 287–303.

Andersen, T. G., & 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., & Diebold, F. X. (2010). Parametric and nonparametric volatility measurement. In Y. Aït-Sahalia, & L. P. Hansen (Eds.), *Handbook of financial econometrics: Tools and techniques* (pp. 67–137). Elsevier.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 579–625.

Andersen, T. G., Su, T., Todorov, V., & Zhang, Z. (2024). Intraday periodic volatility curves. *Journal of the American Statistical Association*, 119(546), 1181–1191

Andersen, T. G., Thyrsgaard, M., & Todorov, V. (2021). Recalcitrant betas: Intraday variation in the cross-sectional dispersion of systematic risk. *Quantitative Economics*, 12, 647–682.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

Bai, Z., & Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & 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., & 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., & Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression and correlation in financial economics. *Econometrica*, 72, 885–925.

Bibinger, M., Hautsch, N., Malec, P., & Reiss, M. (2019). Estimating the spot covariation of asset prices—statistical theory and empirical evidence. *Journal of Business and Economic Statistics*, 37(3), 419–435.

Bibinger, M., Jirak, M., & Vetter, M. (2017). Nonparametric change-point analysis of volatility. *Annals of Statistics*, 45, 1542–1578.

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

Cai, T. T., Hu, J., Li, Y., & Zheng, X. (2020). High-dimensional minimum variance portfolio estimation based on high-frequency data. *Journal of Econometrics*, 214, 482–494.

Cai, T. T., & 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., & Tang, C. (2024). Optimal covariance matrix estimation for high-dimensional noise in high-frequency data. *Journal of Econometrics*, 239, Article 105329.

Chen, J., Li, D., & Linton, O. (2019). A new semiparametric estimation approach of large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics*, 212, 155–176.

Chen, X., Xu, M., & Wu, W. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Annals of Statistics*, 41, 2994–3021.

Chen, Z., & Leng, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association*, 111, 1196–1207.

Christensen, K., Kinnebrock, S., & 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., & 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.

Fan, J., Fan, Y., & Lv, J. (2007). Aggregation of nonparametric estimators for volatility matrix. *Journal of Financial Econometrics*, 5, 321–357.

Fan, J., Furger, A., & 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., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall.

Fan, J., Liao, Y., & Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, 39, 3320–3356.

Fan, J., Liao, Y., & 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., & Wang, Y. (2008). Spot volatility estimation for high-frequency data. *Statistics and Its Interface*, 1, 279–288.

Figueroa-López, J. E., & Li, C. (2020). Optimal kernel estimation of spot volatility of stochastic differential equations. *Stochastic Processes and Their Applications*, 130, 4693–4720.

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., & Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J., & Protter, P. (2012). *Discretization of processes*. Springer.

Jacod, J., & Todorov, V. (2010). Do price and volatility jump together? *Annals of Applied Probability*, 20, 1425–1469.

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

Kanaya, S., & Kristensen, D. (2016). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory*, 32, 861–916.

Kim, D., Wang, Y., & 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. *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., & Feng, P. (2018). A nonparametric eigenvalue-regularized integrated covariance matrix estimator for asset return data. *Journal of Econometrics*, 206, 226–257.

Lee, S., & Mykland, P. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21, 2535–2563.

Li, J., Todorov, V., & Tauchen, G. (2017). Jump regression. *Econometrica*, 85, 173–195.

Li, J., Todorov, V., & Tauchen, G. (2019). Jump factor models in large cross-sections. *Quantitative Economics*, 10, 419–456.

Li, Q., & Racine, J. (2007). *Nonparametric econometrics*. Princeton University Press.

Li, Z. M., & Linton, O. (2022). A ReMeDI for microstructure noise. *Econometrica*, 90, 367–389.

Liao, Y., & Todorov, V. (2024). Changes in the span of systematic risk exposures. *Quantitative Economics*, 15(3), 817–847.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36, 270–296.

Mykland, P. A., & Zhang, L. (2006). ANOVA for diffusions and Itô processes. *Annals of Statistics*, 34, 1931–1963.

Park, S., Hong, S. Y., & 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., & Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15, 634–658.

Reiß, M., Todorov, V., & Tauchen, G. (2015). Nonparametric test for a constant beta between Itô semi-martingales based on high-frequency data. *Stochastic Processes and Their Applications*, 125, 2955–2988.

Revuz, D., & Yor, M. (1999). *Continuous martingales and Brownian motion*. Grundlehren der Mathematischen Wissenschaften (vol. 293). Springer.

Shephard, N. (2005). *Stochastic volatility: Selected readings*. Oxford University Press.

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

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

Xia, N., & 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., & Linton, O. (2002). A nonparametric prewhitened covariance estimator. *Journal of Time Series Analysis*, 23, 215–250.

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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press

Zhang, L., Mykland, P. A., & 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., & Li, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Annals of Statistics*, 39, 3121–3151.

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

https://doi.org/10.1017/S0266466624000264 Published online by Cambridge University Press