OPTIMAL SPARSE VOLATILITY MATRIX ESTIMATION FOR HIGH-DIMENSIONAL ITÔ PROCESSES WITH MEASUREMENT ERRORS

BY MINJING TAO, YAZHEN WANG[1] AND HARRISON H. ZHOU[2]

University of Wisconsin-Madison, University of Wisconsin-Madison and Yale University

Stochastic processes are often used to model complex scientific problems in fields ranging from biology and finance to engineering and physical science. This paper investigates rate-optimal estimation of the volatility matrix of a high-dimensional Itô process observed with measurement errors at discrete time points. The minimax rate of convergence is established for estimating sparse volatility matrices. By combining the multi-scale and threshold approaches we construct a volatility matrix estimator to achieve the optimal convergence rate. The minimax lower bound is derived by considering a subclass of Itô processes for which the minimax lower bound is obtained through a novel equivalent model of covariance matrix estimation for independent but nonidentically distributed observations and through a delicate construction of the least favorable parameters. In addition, a simulation study was conducted to test the finite sample performance of the optimal estimator, and the simulation results were found to support the established asymptotic theory.

1. Introduction. Modern scientific studies in fields ranging from biology and finance to engineering and physical science often need to model complex dynamic systems where it is essential to incorporate internally or externally originating random fluctuations in the systems [Aït-Sahalia, Mykland and Zhang (2005), Mueschke and Andrews (2006) and Whitmore (1995)]. Continuous-time diffusion processes, or more generally, Itô processes, are frequently employed to model such complex dynamic systems. Data collected in the studies are treated as the processes observed at discrete time points with possible noise contamination. For example, the prices of financial assets are usually modeled by Itô processes, and the price data observed at high-frequencies are contaminated by market microstructure noise. In this paper we investigate estimation of the volatilities of the Itô processes based on noisy data.

Several volatility estimation methods have been developed in the past several years. For estimating a univariate integrated volatility, popular estimators include two-scale realized volatility [Zhang, Mykland and Aït-Sahalia (2005)], multi-scale realized volatility [Zhang (2006) and Fan and Wang (2007)], realized kernel volatility [Barndorff-Nielsen et al. (2008)] and pre-averaging based realized volatility [Jacod et al. (2009)]. For estimating a bivariate integrated co-volatility, common methods are the previous-tick approach [Zhang (2011)], the refresh-time scheme and realized kernel volatility [Barndorff-Nielsen et al. (2011)], the generalized synchronization scheme [Aït-Sahalia, Fan and Xiu (2010)] and the pre-averaging approach [Christensen, Kinnebrock and Podolskij (2010)]. Optimal volatility and co-volatility estimation has been investigated in the parametric or nonparametric setting [Aït-Sahalia, Mykland and Zhang (2005), Bibinger and Reiß (2011), Gloter and Jacod (2001a, 2001b), Reiß (2011) and Xiu (2010)]. These works are for estimating scalar volatilities or volatility matrices of small size. Wang and Zou (2010) and Tao et al. (2011) studied the problem of estimating a large sparse volatility matrix based on noisy high-frequency financial data. Fan, Li and Yu (2012) employed a large volatility matrix estimator based on high-frequency data for portfolio allocation. The large volatility matrix estimation is a high-dimensional extension of the univariate case. It can be also considered as a generalization of large covariance matrix estimation for i.i.d. data to volatility matrix estimation for dependent data with measurement errors. Despite recent progress on volatility matrix estimation, there has been remarkably little fundamental theoretical study on optimal estimation of large volatility matrices. Consistent estimation of large matrices based on high-dimensional data usually requires some sparsity, and the sparsity may naturally result from appropriate formulation of some low-dimensional structures in the high-dimensional data. For example, in large volatility matrix estimation with high-frequency financial data sparsity means that a relatively small number of market factors play a dominate role in driving volatility movements and capturing the market risk. In this paper we establish the optimal rate of convergence for large volatility matrix estimation under various matrix norms over a wide range of classes of sparse volatility matrices. We expect that our work will stimulate further theoretical and methodological research as well as more application orientated study on large volatility matrix estimation.

Specifically we consider the problem of estimating the sparse integrated volatility matrix for a p-dimensional Itô process observed with additive noises at n equally spaced discrete time points. The minimax upper bound is obtained by constructing a new procedure through a combination of the multi-scale and threshold approaches and by studying its risk properties. We first construct a multi-scale volatility matrix estimator and show that its elements obey subGaussian tails with a convergence rate $n^{-1/4}$. Then we threshold the constructed estimator to obtain a threshold volatility matrix estimator and derive its convergence rate. The upper bound depends on n and p through $n^{-1/4}\sqrt{\log p}$.

A key step in obtaining the optimal rate of convergence is the derivation of the minimax lower bound for the high-dimensional Itô process with measurement errors. We succeed in establishing the risk lower bound in three steps. First we select a particular subclass of Itô processes with a zero drift and a constant volatility matrix so that the volatility matrix estimation problem becomes a covariance matrix estimation problem where the observed data are dependent and have measurement errors; second, take a special transformation of the observations to convert the problem into a new covariance matrix estimation problem where the observed data have no measurement errors and are independent but not identically distributed, with covariance matrices equal to the constant volatility matrix plus an identity matrix multiplying by a shrinking factor depending on the sample size n; third, adopt the minimax lower bound technique developed in Cai and Zhou (2012) for sparse covariance matrix estimation based on i.i.d. data to establish a minimax lower bound for independent but nonidentically distributed observations. The minimax lower bound matches the upper bound obtained by the new procedure up to a constant factor, and thus the upper bound is rate-optimal.

The volatility matrix estimation is closely related to large covariance matrix estimation which received lots of attentions recently in the literature. While the covariance matrix plays a key role in statistical analysis, its classic estimation procedures, like the sample covariance matrix estimator, may behave very poorly when the matrix size is comparable to or exceeds the sample size. To overcome the curse of dimensionality, various regularization techniques have been developed for estimation of large covariance matrices in recent years. Wu and Pourahmadi (2003) explored nonparametric estimation of large covariance matrices by local stationarity. Ledoit and Wolf (2004) proposed to boost diagonal elements and downgrade off-diagonal elements of the sample covariance matrix estimator. Huang et al. (2006) used a penalized likelihood method to estimate large covariance matrices. Yuan and Lin (2007) considered large covariance matrix estimation in a Gaussian graph model. Bickel and Levina (2008a, 2008b) developed regularization methods by banding or thresholding the sample covariance matrix estimator when the matrix size is comparable to the sample size. El Karoui (2008) employed a graph model approach to characterize sparsity and investigated consistent estimation of large covariance matrices. Fan, Fan and Lv (2008) utilized factor models for estimating large covariance matrices. Johnstone and Lu (2009) studied consistent estimation of leading principal components in principal component analysis. Lam and Fan (2009) established sparsistency and convergence rates for large covariance matrix estimation. Cai, Zhang and Zhou (2010) and Cai and Zhou (2012) studied minimax estimation of covariance matrices when both sample size and matrix size are allowed to go to infinity and derived optimal convergence rates for estimating decaying or sparse covariance matrices.

The rest of the paper proceeds as follows. Section 2 presents the model and the data and constructs volatility matrix estimators. Section 3 establishes the asymptotic theory under sparsity for the constructed matrix estimators as both sample

size and matrix size go to infinity. Section 4 derives the minimax lower bound for estimating a large sparse volatility matrix and shows that the threshold volatility matrix estimator asymptotically achieves the minimax lower bound. Thus combining results in Sections 3 and 4 together, we establish the optimality for large sparse volatility matrix estimation. Section 5 features a simulation study to illustrate the finite sample performances of the volatility matrix estimators. To facilitate the reading we relegate all proofs to Section 6 and two Appendix sections, where we first provide the main proofs of the theorems in Section 6 and then collect additional technical proofs in the two appendices.

2. Volatility matrix estimation.

2.1. *The model set-up.* Suppose that $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^T$ is an Itô process following the model

$$(1) \qquad d\mathbf{X}(t) = \boldsymbol{\mu}_t \, dt + \boldsymbol{\sigma}_t^T \, d\mathbf{B}_t, \qquad t \in [0, 1],$$

where stochastic processes $\mathbf{X}(t)$, $\mathbf{B}_t$, $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in [0, 1]\}, P)$ with filtration $\mathcal{F}_t$ satisfying the usual conditions, $\mathbf{B}_t$ is a p-dimensional standard Brownian motion with respect to $\mathcal{F}_t$, $\boldsymbol{\mu}_t$ is a p-dimensional drift vector, $\boldsymbol{\sigma}_t$ is a p by p matrix, and $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are assumed to be predictable processes with respect to $\mathcal{F}_t$.

We assume that the continuous-time process $\mathbf{X}(t)$ is observed with measurement errors only at equally spaced discrete time points; that is, the observed discrete data $Y_i(t_\ell)$ obey

$$(2) \qquad Y_i(t_\ell) = X_i(t_\ell) + \varepsilon_i(t_\ell), \qquad i = 1, \ldots, p, t_\ell = \ell/n, \ell = 1, \ldots, n,$$

where $\varepsilon_i(t_\ell)$ are noises with mean zero.

Let $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}_t^T \boldsymbol{\sigma}_t$ be the volatility matrix of $\mathbf{X}(t)$. We are interested in estimating the following integrated volatility matrix of $\mathbf{X}(t)$,

$$\boldsymbol{\Gamma} = (\Gamma_{ij})_{1 \le i, j \le p} = \int_0^1 \boldsymbol{\gamma}(t) \, dt = \int_0^1 \boldsymbol{\sigma}_t^T \boldsymbol{\sigma}_t \, dt$$

based on noisy discrete data $Y_i(t_\ell)$, $i = 1, \ldots, p$, $\ell = 1, \ldots, n$.

2.2. *Estimator.* Let K be an integer and $\lfloor n/K \rfloor$ be the largest integer $\le n/K$. We divide n time points $t_1, \ldots, t_n$ into K nonoverlap groups $\boldsymbol{\tau}^k = \{t_\ell, \ell = k, K + k, 2K + k, \ldots\}$, $k = 1, \ldots, K$. Denote by $|\boldsymbol{\tau}^k|$ the number of time points in $\boldsymbol{\tau}^k$. Obviously, the value of $|\boldsymbol{\tau}^k|$ is either $\lfloor n/K \rfloor$ or $\lfloor n/K \rfloor + 1$. For $k = 1, \ldots, K$, we write the rth time point in $\boldsymbol{\tau}^k$ as $\tau_r^k = t_{(r-1)K+k}$, $r = 1, \ldots, |\boldsymbol{\tau}^k|$. With each $\boldsymbol{\tau}^k$, we define the volatility matrix estimator

$$
\tilde{\Gamma}_{ij}(\boldsymbol{\tau}^k) = \sum_{r=2}^{|\boldsymbol{\tau}^k|} [Y_i(\tau_r^k) - Y_i(\tau_{r-1}^k)][Y_j(\tau_r^k) - Y_j(\tau_{r-1}^k)],
$$

$$(3)$$

$$
\tilde{\boldsymbol{\Gamma}}(\boldsymbol{\tau}^k) = (\tilde{\Gamma}_{ij}(\boldsymbol{\tau}^k))_{1 \le i, j \le p}.
$$

Here in (3), to account for noises in data $Y_i(t_\ell)$, we use $\boldsymbol{\tau}^k$ to subsample the data and define $\tilde{\boldsymbol{\Gamma}}(\boldsymbol{\tau}^k)$. To reduce the noise effect we average K volatility matrix estimators $\tilde{\boldsymbol{\Gamma}}(\boldsymbol{\tau}^k)$ to define one-scale volatility matrix estimator

$$(4) \qquad \tilde{\Gamma}_{ij}^K = \frac{1}{K} \sum_{k=1}^K \tilde{\Gamma}_{ij}(\boldsymbol{\tau}^k), \qquad \tilde{\boldsymbol{\Gamma}}^K = (\tilde{\Gamma}_{ij}^K) = \frac{1}{K} \sum_{k=1}^K \tilde{\boldsymbol{\Gamma}}(\boldsymbol{\tau}^k).$$

Let $N = [cn^{1/2}]$ for some positive constant c, and $K_m = m + N$, $m = 1, \ldots, N$. We use each K_m to define a one-scale volatility matrix estimator $\tilde{\boldsymbol{\Gamma}}^{K_m}$ and then combine them together to form a multi-scale volatility matrix estimator

$$(5) \qquad \tilde{\boldsymbol{\Gamma}} = \sum_{m=1}^N a_m \tilde{\boldsymbol{\Gamma}}^{K_m} + \zeta(\tilde{\boldsymbol{\Gamma}}^{K_1} - \tilde{\boldsymbol{\Gamma}}^{K_N}),$$

where

$$(6) \qquad \zeta = \frac{K_1 K_N}{n(N-1)}, \qquad a_m = \frac{12 K_m(m - N/2 - 1/2)}{N(N^2 - 1)},$$

which satisfy

$$\sum_{m=1}^N a_m = 1, \qquad \sum_{m=1}^N \frac{a_m}{K_m} = 0, \qquad \sum_{m=1}^N |a_m| = 9/2 + o(1).$$

The one-scale matrix estimator in (4) was studied in Wang and Zou (2010), and the multi-scale scheme (5)–(6) in the univariate case was investigated in Zhang (2006).

We threshold $\tilde{\boldsymbol{\Gamma}}$ to obtain our final volatility matrix estimator

$$(7) \qquad \widehat{\boldsymbol{\Gamma}} = (\tilde{\Gamma}_{ij} 1(|\tilde{\Gamma}_{ij}| \geq \varpi)),$$

where ϖ is a threshold value to be specified in Theorem 2.

In the estimation construction we use only time scales corresponding to K_m of order $\sqrt{n}$ to form increments and averages. In Section 3 we will demonstrate that the data at these scales contain essential information for estimating $\boldsymbol{\Gamma}$ and show that $\widehat{\boldsymbol{\Gamma}}$ is asymptotically an optimal estimator of $\boldsymbol{\Gamma}$.

3. Asymptotic theory for volatility matrix estimators.

First we fix notation for our asymptotic analysis. Let $\mathbf{x} = (x_1, \ldots, x_p)^T$ be a p-dimensional vector and $\mathbf{A} = (A_{ij})$ be a p by p matrix, and define their ℓ_d norms

$$\|\mathbf{x}\|_d = \left(\sum_{i=1}^p |x_i|^d \right)^{1/d}, \qquad \|\mathbf{A}\|_d = \sup\{\|\mathbf{A}\mathbf{x}\|_d, \|\mathbf{x}\|_d = 1\}, \qquad 1 \leq d \leq \infty.$$

For the case of matrix, the ℓ_2 norm is called the matrix spectral norm. $\|\mathbf{A}\|_2$ is equal to the square root of the largest eigenvalue of $\mathbf{A}\mathbf{A}^T$,

$$(8) \qquad \|\mathbf{A}\|_1 = \max_{1 \leq j \leq p} \sum_{i=1}^{p} |A_{ij}|, \qquad \|\mathbf{A}\|_\infty = \max_{1 \leq i \leq p} \sum_{j=1}^{p} |A_{ij}|$$

and

$$(9) \qquad \|\mathbf{A}\|_2^2 \leq \|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty.$$

For symmetric $\mathbf{A}$, (8)–(9) imply that $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$, and $\|\mathbf{A}\|_2$ is equal to the largest absolute eigenvalue of $\mathbf{A}$.

Second we state some technical conditions for the asymptotic analysis.

A1. Assume $n^{\beta/2} \leq p \leq \exp(\beta_0 \sqrt{n})$ for some constants $\beta > 1$ and $\beta_0 > 0$, and that $\varepsilon_i(t_\ell)$ and $\mathbf{X}(t)$ in models (1)–(2) are independent. Suppose that $(\varepsilon_1(t_\ell), \ldots, \varepsilon_p(t_\ell))$, $\ell = 1, \ldots, n$, is a strictly stationary M-dependent multivariate time series with mean zero and $\mathrm{Var}[\varepsilon_i(t_\ell)] = \eta_i \leq \kappa^2$, where M is a fixed integer, and κ is a finite positive constant. Assume further that $\varepsilon_i(t_\ell)$ are subGaussian in the sense that there exist constants $\tau_0 > 0$ and $c_0 > 0$ such that for all $x > 0$ and $\mathbf{u} = (u_1, \ldots, u_n)^T$ with $\|\mathbf{u}\|_2 = 1$,

$$(10) \qquad P\big(\big|\big(\varepsilon_i(t_1), \ldots, \varepsilon_i(t_n)\big)\mathbf{u}\big| > x\big) \leq c_0 e^{-x^2/(2\tau_0)}, \qquad i = 1, \ldots, p.$$

A2. Assume that there exist positive constants c_1 and c_2 such that

$$\max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} |\mu_i(t)| \leq c_1, \qquad \max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} \gamma_{ii}(t) \leq c_2.$$

Further we assume with probability one for $t \in [0, 1]$,

$$\gamma_{ii}(t) > 0, i = 1, \ldots, p, \qquad \gamma_{ii}(t) + \gamma_{jj}(t) \pm 2\gamma_{ij}(t) > 0,$$

$$i \neq j, i, j = 1, \ldots, p.$$

A3. Assume that $\boldsymbol{\Gamma}$ is sparse in the sense that

$$(11) \qquad \sum_{j=1}^{p} |\Gamma_{ij}|^q \leq \Psi \pi_n(p), \qquad i = 1, \ldots, p,$$

where Ψ is a positive random variable with finite second moment, $0 \leq q < 1$, and $\pi_n(p)$ is a deterministic function with slow growth in p such as $\log p$.

Condition A1 allows noises to have cross sectional correlations as well as cross temporal correlations. In particular we may have any contemporaneous correlations between $\varepsilon_i(t_\ell)$ and $\varepsilon_j(t_\ell)$ as well as lagged serial auto-correlations for individual noise $\varepsilon_i(\cdot)$ and lagged serial cross-correlations between $\varepsilon_i(\cdot)$ and $\varepsilon_j(\cdot)$ with lags up to M. As in covariance matrix estimation, the subGaussianity (10) is essentially required to obtain an optimal convergence rate depending on p through

$\sqrt{\log p}$. It is obvious that independent normal noises satisfy these assumptions. The constraint $p \geq n^{\beta/2}$ is needed to obtain a high-dimensional minimax lower bound; otherwise the problem will be similar to usual asymptotics with large n but fixed p; $p \leq \exp(\beta_0 \sqrt{n})$ is to ensure the existence of a consistent estimator of $\boldsymbol{\Gamma}$. Condition A2 is to impose proper assumptions on the drift and volatility of the Itô process so that we can obtain subGaussian tails for the quadratic forms of $X_i(t_\ell)$, which together with the subGaussianity (10) are used to derive subGaussian tails for the elements of the volatility matrix estimator $\tilde{\boldsymbol{\Gamma}}$. Condition A3 is a common sparsity assumption required for consistently estimating large matrices [Bickel and Levina (2008b), Cai and Zhou (2012), and Johnstone and Lu (2009)].

The following two theorems establish asymptotic theory for the estimators $\tilde{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\Gamma}}$ defined by (5) and (7), respectively.

THEOREM 1.   *Under models* (1)–(2) *and conditions* A1–A2, *the estimator* $\tilde{\boldsymbol{\Gamma}}$ *in* (5) *satisfies that for* $1 \leq i, j \leq p$ *and positive* x *in a neighbor of* 0,

$$(12) \qquad \mathbb{P}(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| \geq x) \leq \varsigma_1 \exp\{\log n - \sqrt{n}x^2/\varsigma_0\},$$

where ς_0 *and* ς_1 *are positive constants free of* n *and* p.

REMARK 1.   Theorem 1 establishes subGaussian tails for the elements of the matrix estimator $\tilde{\boldsymbol{\Gamma}}$. It is known that, when univariate or bivariate continuous Itô processes are observed with measurement errors at n discrete time points, the optimal convergence rates for estimating a univariate integrated volatility or a bivariate integrated co-volatility are $n^{-1/4}$ [Gloter and Jacod (2001a, 2001b), Reiß (2011), and Xiu (2010)]. The $\sqrt{n}x^2$ factor in the exponent of the tail probability bound on the right-hand side of (12) indicates a $n^{-1/4}$ convergence rate for $\tilde{\Gamma}_{ij} - \Gamma_{ij}$, which matches the optimal convergence rate for the univariate integrated volatility estimation. This is in contrast to sub-optimal convergence rate results in the literature where a $n^{-1/6}$ convergence rate was obtained; see, for example, Fan, Li and Yu (2012), Wang and Zou (2010), Zhang, Mykland and Aït-Sahalia (2005), and Zheng and Li (2011).

THEOREM 2.   *For the threshold estimator* $\widehat{\boldsymbol{\Gamma}}$ *in* (7) *we choose threshold* $\varpi = \hbar n^{-1/4}\sqrt{\log(np)}$ *with any fixed constant* $\hbar \geq 5\sqrt{\varsigma_0}$, *where* ς_0 *is the constant in the exponent of the tail probability bound on the right-hand side of* (12). *Denote by* $\mathcal{P}_q(\pi_n(p))$ *the set of distributions of* $Y_i(t_\ell)$, $i = 1, \ldots, p$, $\ell = 1, \ldots, n$, *from models* (1)–(2) *satisfying conditions* A1–A3. *Then as* $n, p \to \infty$,

$$(13) \qquad \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_2^2 \leq \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_1^2$$

$$\leq C^*[\pi_n(p)(n^{-1/4}\sqrt{\log p})^{1-q}]^2,$$

where C^* *is a constant free of* n *and* p.

REMARK 2. For sparse covariance matrix estimation, Cai and Zhou (2012) has shown that the threshold estimator in Bickel and Levina (2008b) is rate-optimal, and the optimal convergence rate depends on n and p through $n^{-1/2} \times \sqrt{\log p}$. The convergence rate obtained in Theorem 2 depends on the sample size n and the matrix size p through $n^{-1/4}\sqrt{\log p}$. Note that $n^{-1/4}$ is the optimal convergence rate for estimating a univariate integrated volatility or a bivariate integrated co-volatility based on noisy data. Since our estimation problem is a generalization of covariance matrix estimation for i.i.d. data to volatility matrix estimation for an Itô process with measurement errors on one hand and a high-dimensional extension of univariate volatility estimation on the other hand, it is interesting to see that the convergence rate in Theorem 2 is a natural blend of convergence rates in the two cases. Also as Theorem 2 implies that the maximum of the eigenvalue differences between $\widehat{\Gamma}$ and Γ is bounded by $\sqrt{C^*}\pi_n(p)(n^{-1/4}\sqrt{\log p})^{1-q}$. Thus if the eigenvalues of Γ all exceed $\sqrt{C^*}\pi_n(p)(n^{-1/4}\sqrt{\log p})^{1-q}$, asymptotically the eigenvalues of $\widehat{\Gamma}$ are positive, and $\widehat{\Gamma}$ is a positive definite matrix. In particular, if $\pi_n(p)(n^{-1/4}\sqrt{\log p})^{1-q}$ goes to zero as n and p go to infinity, and Γ is positive definite and well conditioned, then $\widehat{\Gamma}$ is asymptotically positive definite and well conditioned. In Section 4 we will establish the minimax lower bound for estimating Γ and show that the convergence rate in Theorem 2 is optimal.

4. Optimal convergence rate. This section establishes the minimax lower bound for estimating Γ under models (1)–(2) and shows that asymptotically $\widehat{\Gamma}$ achieves the lower bound and thus is optimal. We state the minimax lower bound for estimating Γ with $\mathcal{P}_q(\pi_n(p))$ under the matrix spectral norm as follows.

THEOREM 3. *For models* (1)–(2) *satisfying conditions* A1–A3, *if for some constant* $\aleph > 0$,

$$\pi_n(p) \leq \aleph n^{(1-q)/4}/(\log p)^{(3-q)/2}, \tag{14}$$

the minimax risk for estimating Γ *with* $\mathcal{P}_q(\pi_n(p))$ *satisfies that as* $n, p \to \infty$,

$$\inf_{\check{\Gamma}} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\|\check{\Gamma} - \Gamma\|_2^2 \geq C_*\big[\pi_n(p)\big(n^{-1/4}\sqrt{\log p}\big)^{1-q}\big]^2, \tag{15}$$

where C_* *is a positive constant free of* n *and* p, *and the infimum is taken over all estimators* $\check{\Gamma}$ *based on the data* $Y_i(t_\ell), i = 1, \ldots, p, \ell = 1, \ldots, n,$ *from models* (1)–(2).

REMARK 3. Note that the lower bound convergence rate in Theorem 3 matches the convergence rate of the estimator $\widehat{\Gamma}$ obtained in Theorem 2. Combining Theorems 2 and 3 together we conclude that the optimal convergence rate is $\pi_n(p)(n^{-1/4}\sqrt{\log p})^{1-q}$, and the estimator $\widehat{\Gamma}$ in (7) achieves the optimal convergence rate. Moreover, such optimal estimation results hold for any matrix ℓ_d norm

with $1 \leq d \leq \infty$. Indeed, it can be shown that under the conditions of Theorems 2 and 3, we have that as n and p go to infinity,

(16)
$$
\frac{C_*}{4}\big[\pi_n(p)\big(n^{-1/4}\sqrt{\log p}\big)^{1-q}\big]^2
$$
$$
\leq \inf_{\check{\mathbf{\Gamma}}} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_d^2 \leq \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_d^2
$$
$$
\leq C^*\big[\pi_n(p)\big(n^{-1/4}\sqrt{\log p}\big)^{1-q}\big]^2,
$$

where C^* and C_* are constants in Theorems 2 and 3, respectively, $\widehat{\mathbf{\Gamma}}$ is the threshold estimator given by (7) with the threshold value specified in Theorem 2 and the infimum is taken over all estimators $\check{\mathbf{\Gamma}}$ based on the data $Y_i(t_\ell)$, $i = 1, \ldots, p$, $\ell = 1, \ldots, n$, from models (1)–(2).

REMARK 4. Condition (14) is a technical condition that we need to establish the minimax lower bound. It is compatible with conditions A1 and A3 regarding the constraint on n and p as well as the slow growth of $\pi_n(p)$ in the sparsity condition (11).

Models (1)–(2) are complicated nonparametric models, and the observations from the models are dependent and have subGaussian measurement errors. To derive the minimax lower bound for models (1)–(2), we find a special subclass of the models to attain the minimax lower bound of the models. Such an approach is often referred to as the method of hardest subproblem. Since generally a minimax problem has lower bound no larger than any of its subproblems, the mentioned special subclass corresponds to the hardest subproblem and is referred to as the least favorable submodel. We will show in Sections 4.1 and 4.2 that the least favorable submodel for models (1)–(2) can be taken as i.i.d. Gaussian measurement errors $\varepsilon_i(t_\ell)$ and process $\mathbf{X}(t)$ with zero drift and constant volatilities. To establish the minimax lower bound for the least favorable submodel, luckily we are able to find a nice trick in Section 4.1 that transforms the minimax lower bound problem for the least favorable submodel into a new covariance matrix estimation problem with independent but nonidentically distributed observations. Cai and Zhou (2012) have developed an approach combining both Le Cam's method and Assouad's lemma, which are two popular methods to establish minimax lower bounds, to derive the minimax lower bound for estimating a large sparse covariance matrix based on i.i.d. observations. We adopt the approach in Cai and Zhou (2012) to derive the minimax lower bound for the new covariance matrix estimation problem with independent but nonidentically distributed observations, which is stated in Theorem 4 of Section 4.2. The derived minimax lower bound in Theorem 4 corresponds to the least favorable submodel and thus is the minimax lower bound for models (1)–(2). Therefore, we prove Theorem 3.

4.1. *Model transformation.*   We take a subclass of models (1)–(2) as follows. For the Itô processes $\mathbf{X}(t)$ we let $\boldsymbol{\mu}_t = 0$ and $\boldsymbol{\sigma}_t$ be a constant matrix $\boldsymbol{\sigma}$; for the noises we let $\varepsilon_i(t_\ell)$, $i = 1, \ldots, p$, $\ell = 1, \ldots, n$, be i.i.d. random variables with $N(0, \kappa^2)$ distribution, where $\kappa > 0$ is specified in condition A1. Then $\boldsymbol{\Gamma} = (\Gamma_{ij}) = \boldsymbol{\sigma}^T \boldsymbol{\sigma}$, and the sparsity condition (11) becomes

$$(17) \qquad \sum_{j=1}^{p} |\Gamma_{ij}|^q \leq c_3 \pi_n(p),$$

where $c_3 = E(\Psi)$ and Ψ is given by (11).

Let $\mathbf{Y}_l = (Y_1(t_l), \ldots, Y_p(t_l))^T$, and $\boldsymbol{\varepsilon}_l = (\varepsilon_1(t_l), \ldots, \varepsilon_p(t_l))^T$. Then models (1)–(2) become

$$(18) \qquad\qquad \mathbf{Y}_l = \boldsymbol{\sigma} \mathbf{B}_{t_l} + \boldsymbol{\varepsilon}_l, \qquad l = 1, \ldots, n, t_l = l/n$$

and $\boldsymbol{\varepsilon}_l \sim N(0, \kappa^2 I_p)$. As $\mathbf{Y}_l$ are dependent, we take differences in (18) and obtain

$$(19) \qquad \mathbf{Y}_l - \mathbf{Y}_{l-1} = \boldsymbol{\sigma}(\mathbf{B}_{t_l} - \mathbf{B}_{t_{l-1}}) + \boldsymbol{\varepsilon}_l - \boldsymbol{\varepsilon}_{l-1}, \qquad l = 1, \ldots, n,$$

here $\mathbf{Y}_0 = \boldsymbol{\varepsilon}_0 \sim N(0, \kappa^2 \mathbf{I}_p)$. For matrix $(\boldsymbol{\varepsilon}_l - \boldsymbol{\varepsilon}_{l-1}, 1 \leq l \leq n) = (\varepsilon_i(t_l) - \varepsilon_i(t_{l-1}), 1 \leq i \leq p, 1 \leq l \leq n)$, its elements are independent at different rows but correlated at the same rows. At the ith row, elements $\varepsilon_i(t_l) - \varepsilon_i(t_{l-1})$, $l = 1, \ldots, n$, have covariance matrix $\kappa^2 \boldsymbol{\Upsilon}$, where $\boldsymbol{\Upsilon}$ is a $n \times n$ tridiagonal matrix with 2 along diagonal entries, -1 next to diagonal entries and 0 elsewhere. $\boldsymbol{\Upsilon}$ is a Toeplitz matrix [Wilkinson (1988)] that can be diagonalized as follows:

$$(20) \qquad\qquad \boldsymbol{\Upsilon} = \mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^T, \qquad \boldsymbol{\Phi} = \mathrm{diag}(\varphi_1, \ldots, \varphi_n),$$

where φ_l are eigenvalues with expressions

$$(21) \qquad\qquad \varphi_l = 4\sin^2\left[\frac{\pi l}{2(n+1)}\right], \qquad l = 1, \ldots, n,$$

and $\mathbf{Q}$ is an orthogonal matrix formed by the eigenvectors of $\boldsymbol{\Upsilon}$. Using (20) we transform the ith row of the matrix $(\boldsymbol{\varepsilon}_l - \boldsymbol{\varepsilon}_{l-1}, 1 \leq l \leq n)$ by $\mathbf{Q}$, and obtain

$$\mathrm{Var}\big[(\varepsilon_i(t_1) - \varepsilon_i(t_0), \ldots, \varepsilon_i(t_n) - \varepsilon_i(t_{n-1}))\mathbf{Q}\big] = \kappa^2 \mathbf{Q}^T \boldsymbol{\Upsilon} \mathbf{Q} = \kappa^2 \boldsymbol{\Phi}.$$

For $i = 1, \ldots, p$, let

$$(e_{i1}, \ldots, e_{in}) = \big(\sqrt{n}[\varepsilon_i(t_1) - \varepsilon_i(t_0)], \ldots, \sqrt{n}[\varepsilon_i(t_n) - \varepsilon_i(t_{n-1})]\big)\mathbf{Q},$$

$$(u_{i1}, \ldots, u_{in}) = \big(\sqrt{n}[Y_i(t_1) - Y_i(t_0)], \ldots, \sqrt{n}[Y_i(t_n) - Y_i(t_{n-1})]\big)\mathbf{Q},$$

$$(v_{i1}, \ldots, v_{in}) = \big(\sqrt{n}[B_i(t_1) - B_i(t_0)], \ldots, \sqrt{n}[B_i(t_n) - B_i(t_{n-1})]\big)\mathbf{Q}.$$

Then as $\mathbf{Q}$ diagonalizes $\boldsymbol{\Upsilon}$, e_{il} are independent, with $e_{il} \sim N(0, n\kappa^2\varphi_l)$; because $B_i(t_l) - B_i(t_{l-1})$ are i.i.d. normal random variables with mean zero and variance $1/n$, and $\mathbf{Q}$ is orthogonal, v_{il} are i.i.d. standard normal random variables.

Put (19) in a matrix form and right multiply by $\sqrt{n}\mathbf{Q}$ on both sides to obtain

$$(u_{il}) = \boldsymbol{\sigma}(v_{il}) + (e_{il}).$$

Denote by $\mathbf{U}_l$, $\mathbf{V}_l$ and $\mathbf{e}_l$ the column vectors of the matrices (u_{il}), (v_{il}) and (e_{il}), respectively. Then the above matrix equation is equivalent to

$$(22) \qquad\qquad \mathbf{U}_l = \boldsymbol{\sigma}\mathbf{V}_l + \mathbf{e}_l, \qquad l = 1, \ldots, n,$$

where $\mathbf{e}_l \sim N(0, \kappa^2 n\varphi_l \mathbf{I}_p)$ and $\mathbf{V}_l \sim N(0, \mathbf{I}_p)$.

From (22) we have that the data transformed random vectors $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are independent with $\mathbf{U}_l \sim N(0, \boldsymbol{\Gamma} + (a_l - 1)\mathbf{I}_p)$, where $a_l = 1 + \kappa^2 n\varphi_l$ with $0 < \kappa < \infty$.

4.2. *Lower bound.* We convert the minimax lower bound problem stated in Theorem 3 into a much simpler problem of estimating $\boldsymbol{\Gamma}$ based on the observations $\mathbf{U}_1, \ldots, \mathbf{U}_n$ from model (22), where $\boldsymbol{\Gamma}$ are constant matrices satisfying (17) and $\|\boldsymbol{\Gamma}\|_2 \leq \tau$ for some constant $\tau > 0$. We denote the new minimax estimation problem by $\mathcal{Q}_q(\pi_n(p))$, and the theorem below derives its minimax lower bound.

THEOREM 4. *Assume $p \geq n^{\beta/2}$ for some $\beta > 1$. If $\pi_n(p)$ obeys (14), the minimax risk for estimating matrix $\boldsymbol{\Gamma}$ with $\mathcal{Q}_q(\pi_n(p))$ satisfies that as $n, p \to \infty$,*

$$(23) \qquad \inf_{\check{\boldsymbol{\Gamma}}} \sup_{\mathcal{Q}_q(\pi_n(p))} \mathbb{E}\|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_2^2 \geq C_*\big[\pi_n(p)\big(n^{-1/4}\sqrt{\log p}\,\big)^{1-q}\big]^2,$$

where C_* is a positive constant free of n and p, and the infimum is taken over all estimators $\check{\boldsymbol{\Gamma}}$ based on the observations $\mathbf{U}_1, \ldots, \mathbf{U}_n$ from model (22).

REMARK 5. As we discussed in Remarks 1 and 2 in Section 3, due to noise contamination, the optimal convergence rate depends on sample size through $n^{-1/4}$, instead of $n^{-1/2}$ for covariance matrix estimation. For the univariate case, discrete sine transform was used to construct a realized volatility estimator [Aït-Sahalia, Mykland and Zhang (2005) and Curci and Corsi (2012)] and reveal some intrinsic insight into how the $n^{-1/4}$ convergence rate is obtained [Munk and Schmidt-Hieber (2010)]. The similar insight for the high-dimensional case can be seen from the transformation in Section 4.1, which converts model (19) with noisy data into model (22) where the independent random vector $\mathbf{U}_l$ follows a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Gamma} + \kappa^2 n\varphi_l \mathbf{I}_p$, $l = 1, \ldots, n$. The transformation via orthogonal matrix $\mathbf{Q}$, which diagonalizes Toeplitz matrix $\boldsymbol{\Upsilon}$ and is equal to $(\sin(\ell r\pi/(n + 1)), 1 \leq \ell, r \leq n)$ normalized by $\sqrt{2/(n + 1)}$ [see Salkuyeh (2006)], corresponds to a discrete sine transform, with (22) in frequency domain and $\mathbf{U}_l \sim N(0, \boldsymbol{\Gamma} + \kappa^2 n\varphi_l \mathbf{I}_p)$ corresponding to the discrete sine transform of the data at frequency $l\pi/(n + 1)$. By comparing the order of $n\varphi_l$, we derive that only at those frequencies with l up to $\sqrt{n}$, the transformed

data $\mathbf{U}_l$ are informative for estimating $\mathbf{\Gamma}$, and we use these $[\sqrt{n}]$ number of $\mathbf{U}_l$ to estimate $\mathbf{\Gamma}$ and obtain $(\sqrt{n})^{-1/2} = n^{-1/4}$ convergence rate. In fact, we have seen the phenomenon in Section 2.1 where the N scales used in the construction of $\tilde{\mathbf{\Gamma}}$ in (5) correspond to K_m, with both N and K_m of order $\sqrt{n}$.

5. A simulation study. A simulation study was conducted to compare the finite sample performances of the MSRVM estimator in (5) and the threshold MSRVM estimator in (7) with those of the ARVM estimator and the threshold ARVM estimator introduced in Wang and Zou (2010). We generated $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^T$ at discrete time points $t_\ell = \ell/n$, $\ell = 1, \ldots, n$, from model (1) with $\boldsymbol{\mu}_t = 0$ by the Euler scheme, where univariate standard Brownian motions were stimulated by the normalized partial sums of independent standard normal random variables, $\boldsymbol{\sigma}_{t_\ell}$ was taken to be a Cholesky decomposition of

$$\boldsymbol{\gamma}(t_\ell) = (\gamma_{ij}(t_\ell)), \qquad \gamma_{ij}(t_\ell) = \sqrt{\gamma_{ii}(t_\ell)\gamma_{jj}(t_\ell)} \varrho^{|i-j|},$$

ϱ was independently generated from a uniform distribution on $[0.47, 0.53]$, $(\gamma_{ii}(t_1), \ldots, \gamma_{ii}(t_n))$, $i = 1, \ldots, p$, were independently drawn from a geometric Ornstein–Uhlenbeck process satisfying $d \log \gamma_{ii}(t) = 6[0.5 - \log \gamma_{ii}(t)] \, dt + dW_i(t)$ and $W_i(t)$ are independent one-dimensional standard Brownian motions that are independent of $\mathbf{B}_t$ in model (1). We computed $\mathbf{\Gamma}$ by the average of $\boldsymbol{\gamma}(t_1), \ldots, \boldsymbol{\gamma}(t_n)$. We simulated noises $\varepsilon_i(t_\ell)$ independently from a normal distribution with mean 0 and standard deviation $\theta \sqrt{\Gamma_{ii}}$, $i = 1, \ldots, p$, where θ is the relative noise level ranging from 0 to 0.7. Finally data $Y_i(t_\ell)$ were obtained by adding the simulated $\varepsilon_i(t_\ell)$ to the generated $X_i(t_\ell)$ according to model (2). Using the simulated data $Y_i(t_\ell)$ we computed the MSRVM estimator and the threshold MSRVM estimator as well as the ARVM estimator and the threshold ARVM estimator. In the simulation study we took $n = 200$ and $p = 100$. We repeated the whole simulation procedure 200 times. For a given matrix estimator $\check{\mathbf{\Gamma}}$, a relative matrix spectral norm error $\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_2 / \|\mathbf{\Gamma}\|_2$ was used to measure its performance. We evaluated the mean relative matrix spectral norm error (MRE) by the average of the relative matrix spectral norm errors over the 200 repetitions. As in Wang and Zou (2010) we selected tuning parameters like threshold of the estimators by minimizing the respective MREs.

Figure 1 is the plots of MRE versus relative noise level θ for the MSRVM, ARVM, threshold MSRVM and threshold ARVM estimators. The basic findings are that while the MREs of the threshold MSRVM and threshold ARVM estimators are comparable at low relative noise levels, the threshold MSRVM estimator has smaller MRE than the threshold ARVM estimator at high relative noise levels; regardless of relative noise levels, the threshold MSRVM and threshold ARVM estimators have significantly smaller MREs than the MSRVM and ARVM estimators. The simulation results support the theoretical conclusions that the threshold procedure is needed for constructing consistent estimators of $\mathbf{\Gamma}$, and the threshold
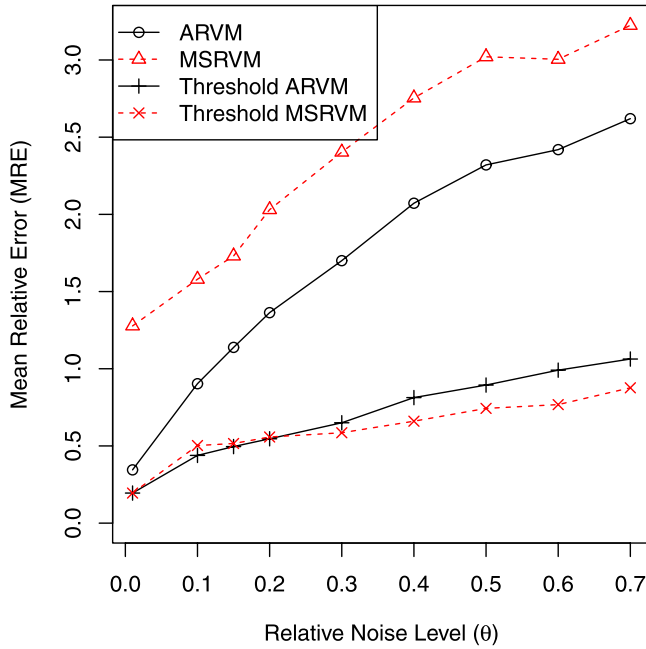
FIG. 1. *The MRE plots of the four estimators for $n = 200$ and $p = 100$.*

MSRVM estimator is asymptotically optimal, while the threshold ARVM estimator is suboptimal.

We point out that it is important to have a data-driven choice of tuning parameters for volatility matrix estimator defined in (7). This is largely an open issue. We briefly describe an approach for developing a data-dependent selection of the tuning parameters as follows. For data $\{Y_i(t_\ell), i = 1, \ldots, p, \ell = 1, \ldots, n\}$ observed from models (1)–(2), we may divide the whole data time interval into L subintervals $I_1, \ldots, I_L$, and partition data $Y_i(t_\ell)$ into L subsamples $\{Y_i(t_\ell), i = 1, \ldots, p, t_\ell \in I_k\}$, $k = 1, \ldots, L$, over the L corresponding time periods. To estimate integrated volatility $\int_{I_k} \boldsymbol{\gamma}(t) \, dt / |I_k|$ over the kth period, according to the procedure described in Section 2.2, we use the kth subsample to construct volatility matrix estimator, which is denoted by $\widehat{\Gamma}_k(N, \varpi)$ to emphasize its dependence on N and ϖ, where $|I_k|$ denotes the length of I_k, ϖ is a threshold value and N is an integer that specifies scales used in the volatility matrix estimator given by (7). We predict one period ahead volatility matrix estimator $\widehat{\Gamma}_{k+1}(N, \varpi)$ by current period volatility matrix estimator $\widehat{\Gamma}_k(N, \varpi)$ and compute the predication error. We minimize the sum of the spectral norms of the predication errors to select N and ϖ. For example, we often have high-frequency financial data over many days, and it is natural to use data in each day to estimate the integrated volatility matrix over the corresponding day. We predict one day ahead daily volatility matrix estimator

by current daily volatility matrix estimator and compute the predication error. The tuning parameters are then selected by minimizing the sum of the spectral norms of the prediction errors.

6. Proofs. Denote by C's generic constants whose values are free of n and p and may change from appearance to appearance. Let $u \vee v$ and $u \wedge v$ be the maximum and minimum of u and v, respectively. For two sequences $u_{n,p}$ and $v_{n,p}$ we write $u_{n,p} \asymp v_{n,p}$ if there exist positive constants C_1 and C_2 free of n and p such that $C_1 \leq u_{n,p}/v_{n,p} \leq C_2$. Without loss of generality we take $N = [n^{1/2}]$ in the construction of $\tilde{\Gamma}$ given by (5) in Section 2.2.

6.1. *Proofs of Theorems* 1 *and* 2. Let

$$\mathbf{Y}_r^{k_m} = \big(Y_1(\tau_r^{k_m}), \dots, Y_p(\tau_r^{k_m})\big)^T,$$

$$\mathbf{X}_r^{k_m} = \big(X_1(\tau_r^{k_m}), \dots, X_p(\tau_r^{k_m})\big)^T,$$

$$\boldsymbol{\varepsilon}_r^{k_m} = \big(\varepsilon_1(\tau_r^{k_m}), \dots, \varepsilon_p(\tau_r^{k_m})\big)^T,$$

which are random vectors corresponding to the data, the Itô process and the noises at the time point $\tau_r^{k_m}$, $r = 1, \dots, |\tau^{k_m}|$, $k_m = 1, \dots, K_m$, and $m = 1, \dots, N$. Note that we choose index k_m to specify that the analyses are associated with the study of Γ^{K_m} here and below. We decompose $\tilde{\Gamma}^{K_m}$ defined in (4) as follows:

$$
\begin{aligned}
\tilde{\Gamma}^{K_m} &= \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} \big(\mathbf{Y}_r^{k_m} - \mathbf{Y}_{r-1}^{k_m}\big)\big(\mathbf{Y}_r^{k_m} - \mathbf{Y}_{r-1}^{k_m}\big)^T \\
&= \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} \big(\mathbf{X}_r^{k_m} - \mathbf{X}_{r-1}^{k_m} + \boldsymbol{\varepsilon}_r^{k_m} - \boldsymbol{\varepsilon}_{r-1}^{k_m}\big) \\
&\qquad\qquad \times \big(\mathbf{X}_r^{k_m} - \mathbf{X}_{r-1}^{k_m} + \boldsymbol{\varepsilon}_r^{k_m} - \boldsymbol{\varepsilon}_{r-1}^{k_m}\big)^T \\
&= \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} \big\{\big(\mathbf{X}_r^{k_m} - \mathbf{X}_{r-1}^{k_m}\big)\big(\mathbf{X}_r^{k_m} - \mathbf{X}_{r-1}^{k_m}\big)^T
\end{aligned}
$$

$$(24)$$

$$
\begin{aligned}
&\qquad\qquad + \big(\boldsymbol{\varepsilon}_r^{k_m} - \boldsymbol{\varepsilon}_{r-1}^{k_m}\big)\big(\boldsymbol{\varepsilon}_r^{k_m} - \boldsymbol{\varepsilon}_{r-1}^{k_m}\big)^T \\
&\qquad\qquad + \big(\mathbf{X}_r^{k_m} - \mathbf{X}_{r-1}^{k_m}\big)\big(\boldsymbol{\varepsilon}_r^{k_m} - \boldsymbol{\varepsilon}_{r-1}^{k_m}\big)^T \\
&\qquad\qquad + \big(\boldsymbol{\varepsilon}_r^{k_m} - \boldsymbol{\varepsilon}_{r-1}^{k_m}\big)\big(\mathbf{X}_r^{k_m} - \mathbf{X}_{r-1}^{k_m}\big)^T \big\} \\
&\equiv \mathbf{V}^{K_m} + \mathbf{G}^{K_m}(1) + \mathbf{G}^{K_m}(2) + \mathbf{G}^{K_m}(3),
\end{aligned}
$$

and thus from (5) we obtain the corresponding decomposition for $\tilde{\mathbf{\Gamma}}$,

$$\tilde{\mathbf{\Gamma}} = \sum_{m=1}^{N} a_m \mathbf{V}^{K_m} + \zeta (\mathbf{V}^{K_1} - \mathbf{V}^{K_N})$$

$$(25) \qquad + \sum_{r=1}^{3} \left[\sum_{m=1}^{N} a_m \mathbf{G}^{K_m}(r) + \zeta (\mathbf{G}^{K_1}(r) - \mathbf{G}^{K_N}(r)) \right]$$

$$\equiv \mathbf{V} + \mathbf{G}(1) + \mathbf{G}(2) + \mathbf{G}(3),$$

where the $\mathbf{V}^{k_m}$ and $\mathbf{V}$ terms are associated with the process $\mathbf{X}(t)$ only, the $\mathbf{G}^{K_m}(1)$ and $\mathbf{G}(1)$ terms are related to the noises $\varepsilon_i(t_\ell)$ only and the terms denoted by $\mathbf{G}^{K_m}(2)$, $\mathbf{G}^{K_m}(3)$, $\mathbf{G}(2)$ and $\mathbf{G}(3)$ depend on both $\mathbf{X}(t)$ and $\varepsilon_i(t_\ell)$.

Now we may heuristically explain the basic ideas for proving Theorems 1 and 2 as follows. With the expression (25) we prove the tail probability result for $\tilde{\mathbf{\Gamma}}$ in Theorem 1 by establishing tail probabilities for these $\mathbf{V}$ and $\mathbf{G}$ terms in the following three propositions whose proofs will be given in Appendix I.

PROPOSITION 5. *Under the assumptions of Theorem* 1, *we have for* $1 \leq i, j \leq p$ *and positive d in a neighbor of* 0,

$$P(|V_{ij} - \Gamma_{ij}| \geq d) \leq C_1 n \exp\{-\sqrt{n}d^2/C_2\}.$$

PROPOSITION 6. *Under the assumptions of Theorem* 1, *we have for* $1 \leq i, j \leq p$ *and positive d in a neighbor of* 0,

$$P(|G_{ij}(2)| \geq d) \leq C_1 n \exp\{-\sqrt{n}d^2/C_2\},$$

$$P(|G_{ij}(3)| \geq d) \leq C_1 n \exp\{-\sqrt{n}d^2/C_2\}.$$

PROPOSITION 7. *Under the assumptions of Theorem* 1, *we have for* $1 \leq i, j \leq p$ *and positive d in a neighbor of* 0,

$$P(|G_{ij}(1)| \geq d) \leq C_1 \sqrt{n} \exp\{-\sqrt{n}d^2/C_2\}.$$

Because V_{ij} are quadratic forms in the process $\mathbf{X}(t_\ell)$ only, we derive their tail probability in Proposition 5 from the boundedness of the drift and volatility in condition A2; as $G_{ij}(1)$ are quadratic forms in the noises $\varepsilon_i(t_\ell)$ only, we establish the tail probability of $G_{ij}(1)$ in Proposition 7 from the subGaussianity of $\varepsilon_i(t_\ell)$ imposed by condition A1; $G_{ij}(2)$ and $G_{ij}(3)$ are bilinear forms in $\mathbf{X}(t_\ell)$ and $\varepsilon_i(t_\ell)$, thus we obtain the tail probabilities for $G_{ij}(2)$ and $G_{ij}(3)$ in Proposition 6 from the subGaussian tails of $\varepsilon_i(t_\ell)$ and V_{ij} as well as the independence between $\varepsilon_i(t_\ell)$ and $\mathbf{X}(t)$ given by condition A1. Since $\hat{\mathbf{\Gamma}}$ is the matrix estimator obtained by thresholding $\tilde{\mathbf{\Gamma}}$, we use the tail probability result in Theorem 1 and the sparsity of $\mathbf{\Gamma}$ to analyze $\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}$ and control its matrix norm for proving Theorem 2.

PROOF OF THEOREM 1.   From (25) we have

$$P\big(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| \geq x\big) \leq P\big(|V_{ij} - \Gamma_{ij}| \geq x/4\big) + \sum_{r=1}^{3} P\big(|G_{ij}(r)| \geq x/4\big),$$

and thus the theorem is a consequence of Propositions 5–7.   □

PROOF OF THEOREM 2.   Define

$$A_{ij} = \big\{|\widehat{\Gamma}_{ij} - \Gamma_{ij}| \leq 2\min\{|\Gamma_{ij}|, \varpi\}\big\}, \qquad D_{ij} = (\widehat{\Gamma}_{ij} - \Gamma_{ij})1(A_{ij}^c),$$
$$\mathbf{D} = (D_{ij})_{1 \leq i, j \leq p}.$$

As the matrix norm of a symmetric matrix is bounded by its ℓ_1-norm, then

$$(26) \qquad E\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_2^2 \leq E\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_1^2 \leq 2E\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma} - \mathbf{D}\|_1^2 + 2E\|\mathbf{D}\|_1^2.$$

We can bound $E\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma} - \mathbf{D}\|_1^2$ as follows:

$$\begin{aligned} E\|\widehat{\mathbf{\Gamma}} &- \mathbf{\Gamma} - \mathbf{D}\|_1^2 \\ &= E\left[\max_{1 \leq j \leq p} \sum_{i=1}^{p} |\widehat{\Gamma}_{ij} - \Gamma_{ij}|1(|\widehat{\Gamma}_{ij} - \Gamma_{ij}| \leq 2\min\{|\Gamma_{ij}|, \varpi\})\right]^2 \\ &\leq E\left[\max_{1 \leq j \leq p} \sum_{i=1}^{p} 2|\Gamma_{ij}|1(|\Gamma_{ij}| < \varpi)\right]^2 \\ &\quad + E\left[\max_{1 \leq j \leq p} \sum_{i=1}^{p} 2\varpi 1(|\Gamma_{ij}| \geq \varpi)\right]^2 \\ &\leq 8E[\Psi^2]\pi_n^2(p)\varpi^{2(1-q)} \leq C\pi_n^2(p)\big(n^{-1/4}\sqrt{\log p}\big)^{2-2q}, \end{aligned}$$

where the second inequality is due to the fact that the sparsity of $\mathbf{\Gamma}$ implies

$$\max_{1 \leq j \leq p} \sum_{i=1}^{p} 1(|\Gamma_{ij}| \geq \varpi) \leq \Psi\pi_n(p)\varpi^{-q},$$

$$\max_{1 \leq j \leq p} \sum_{i=1}^{p} |\Gamma_{ij}|1(|\Gamma_{ij}| < \varpi) \leq \Psi\pi_n(p)\varpi^{1-q},$$

which are the respective bounds on the number of those entries on each row with absolute values larger than or equal to ϖ and the sum of those absolute entries on each row with magnitudes less than ϖ; see Lemma 1 in Wang and Zou (2010). The rest of the proof is to show that $E\|\mathbf{D}\|_1^2 = O(n^{-2})$, a negligible term. Indeed,

the threshold rule indicates that $\widehat{\Gamma}_{ij} = 0$ if $|\tilde{\Gamma}_{ij}| < \varpi$ and $\widehat{\Gamma}_{ij} = \tilde{\Gamma}_{ij}$ if $|\tilde{\Gamma}_{ij}| \geq \varpi$, thus

$$E\|\mathbf{D}\|_1^2 \leq p \sum_{i,j=1}^p E\big[|\Gamma_{ij}|^2 \mathbf{1}(|\Gamma_{ij}| > 2\min\{|\Gamma_{ij}|, \varpi\}) \mathbf{1}(\widehat{\Gamma}_{ij} = 0)\big]$$

$$+ p \sum_{i,j}^p E\big[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 \mathbf{1}(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > 2\min\{|\Gamma_{ij}|, \varpi\}) \mathbf{1}(\widehat{\Gamma}_{ij} = \tilde{\Gamma}_{ij})\big]$$

$$\equiv I_1 + I_2.$$

For term I_1, we have

$$I_1 \leq p \sum_{i,j=1}^p E\big[|\Gamma_{ij}|^2 \mathbf{1}(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi)\big] \leq Cp \sum_{i,j=1}^p P(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi)$$

$$\leq Cp^3 \exp\{\log n - \sqrt{n}\varpi^2/\varsigma_0\} \leq Cn^{-2},$$

where the third inequality is from Theorem 1, and the last inequality is due to $\varpi = \hbar n^{-1/4}\sqrt{\log(np)}$ with $\hbar^2/\varsigma_0 > 4$.

On the other hand, we can bound term I_2 as follows:

$$I_2 \leq p \sum_{i,j=1}^p E\big[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 \mathbf{1}(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi)\big]$$

$$+ p \sum_{i,j=1}^p E\big[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 \mathbf{1}(|\Gamma_{ij}| < \varpi/2, |\tilde{\Gamma}_{ij}| \geq \varpi)\big]$$

$$\leq 2p \sum_{i,j=1}^p E\big[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 \mathbf{1}(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi/2)\big]$$

$$\leq 2p \sum_{i,j=1}^p \big\{E\big[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^4\big] P(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi/2)\big\}^{1/2}$$

$$\leq Cp^3 \exp\{\log n/2 - \sqrt{n}\varpi^2/(8\varsigma_0)\} \leq Cn^{-2},$$

where the third inequality is due to Hölder's inequality, the fourth inequality is from Theorem 1 and

$$(27) \qquad\qquad \max_{1 \leq i,j \leq p} E\big[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^4\big] \leq C$$

and the last inequality is due to the fact that $\varpi = \hbar n^{-1/4}\sqrt{\log(np)}$ with $\hbar^2/(8\varsigma_0) > 3$.

To complete the proof we need to show (27). As in Zhang, Mykland and Aït-Sahalia (2005), we adjust $\tilde{\mathbf{\Gamma}}^{K_m}$ to account for the noise variances. Let

$$(28) \qquad \tilde{\boldsymbol{\eta}} = \mathrm{diag}(\tilde{\eta}_1, \ldots, \tilde{\eta}_p), \qquad \tilde{\eta}_i = \frac{1}{2n} \sum_{\ell=2}^{n} [Y_i(t_\ell) - Y_i(t_{\ell-1})]^2,$$

and define

$$(29) \qquad \tilde{\mathbf{\Gamma}}^{*K_m} = \tilde{\mathbf{\Gamma}}^{K_m} - 2\frac{n - K_m + 1}{K_m} \tilde{\boldsymbol{\eta}},$$

which are the average realized volatility matrix (ARVM) estimators where the convergence rates for any finite moments of $\tilde{\Gamma}_{ij}^{*K_m} - \Gamma_{ij}$ are derived in Wang and Zou [(2010), Theorem 1]. Applying Theorem 1 of Wang and Zou (2010) to the fourth moment of $\tilde{\Gamma}_{ij}^{*K_m} - \Gamma_{ij}$, we have for $1 \le i, j \le p$ and $1 \le m \le N$,

$$(30) \begin{aligned} &E\big(|\tilde{\Gamma}_{ij}^{*K_m} - \Gamma_{ij}|^4\big) \\ &\qquad \le C\big[(K_m n^{-1/2})^{-4} + K_m^{-2} + (n/K_m)^{-2} + K_m^{-4} + n^{-2}\big] \le C. \end{aligned}$$

From (5), (6) and (29) together with simple algebraic manipulations we can express $\tilde{\mathbf{\Gamma}}$ by $\tilde{\mathbf{\Gamma}}^{*K_m}$ as follows:

$$\tilde{\mathbf{\Gamma}} = \sum_{m=1}^{N} a_m \tilde{\mathbf{\Gamma}}^{*K_m} + \zeta\big(\tilde{\mathbf{\Gamma}}^{*K_1} - \tilde{\mathbf{\Gamma}}^{*K_N}\big),$$

and thus

$$(31) \qquad \tilde{\mathbf{\Gamma}} - \mathbf{\Gamma} = \sum_{m=1}^{N} a_m\big(\tilde{\mathbf{\Gamma}}^{*K_m} - \mathbf{\Gamma}\big) + \zeta\big[(\tilde{\mathbf{\Gamma}}^{*K_1} - \mathbf{\Gamma}) - (\tilde{\mathbf{\Gamma}}^{*K_N} - \mathbf{\Gamma})\big].$$

Combining (30) and (31) and using (6) we conclude for $1 \le i, j \le p$,

$$\begin{aligned} &E\big[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^4\big] \\ &\qquad \le (N+2)^3 \bigg[\sum_{m=1}^{N} a_m^4 E\big(|\tilde{\Gamma}_{ij}^{*K_m} - \Gamma_{ij}|^4\big) \\ &\qquad\qquad\qquad + \zeta^4 E\big(|\tilde{\mathbf{\Gamma}}^{*K_1} - \Gamma_{ij}|^4 + |\tilde{\mathbf{\Gamma}}^{*K_N} - \Gamma_{ij}|^4\big)\bigg] \\ &\qquad \le C. \hspace{8cm} \square \end{aligned}$$

6.2. *Proofs of Theorems 3 and 4.* Section 4.1 shows that Theorem 3 is a consequence of Theorem 4. The proof of Theorem 4 is similar to but much more involved than the proof of Theorem 2 in Cai and Zhou (2012) which considered only i.i.d. observations. It contains four major steps. In the first step we construct in

detail a finite subset $\mathcal{F}_*$ of the parameter space $\mathcal{G}_q(\pi_n(p))$ in the minimax problem $\mathcal{Q}_q(\pi_n(p))$ such that the difficulty of estimation over $\mathcal{F}_*$ is essentially the same as that of estimation over $\mathcal{G}_q(\pi_n(p))$, where $\mathcal{G}_q(\pi_n(p))$ is the class of constant matrices $\mathbf{\Gamma}$ satisfying (17) and $\|\mathbf{\Gamma}\|_2 \leq \tau$ for constant $\tau > 0$. The second step applies the lower bound argument in Cai and Zhou [(2012), Lemma 3] to the carefully constructed parameter set $\mathcal{F}_*$. In the third step we calculate the factor α defined in (40) below and the total variation affinity between two average of products of n independent but nonidentically distributed multivariate normals. The final step combines together the results in steps 2 and 3 to obtain the minimax lower bound.

Step 1: *Construct parameter set* $\mathcal{F}_*$. Set $r = \lceil p/2 \rceil$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, and let B be the collection of all row vectors $b = (v_j)_{1 \leq j \leq p}$ such that $v_j = 0$ for $1 \leq j \leq p - r$ and $v_j = 0$ or 1 for $p - r + 1 \leq j \leq p$ under the constraint $\|b\|_0 = k$ (to be specified later). Each element $\lambda = (b_1, \ldots, b_r) \in B^r$ is treated as an $r \times p$ matrix with the ith row of λ equal to b_i. Let $\Delta = \{0, 1\}^r$. Define $\Lambda \subset B^r$ to be the set of all elements in B^r such that each column sum is less than or equal to $2k$. For each $b \in B$ and each $1 \leq m \leq r$, define a $p \times p$ symmetric matrix $A_m(b)$ by making the mth row of $A_m(b)$ equal to b, mth column equal to b^T and the rest of the entries 0. Then each component λ_i of $\lambda = (\lambda_1, \ldots, \lambda_r) \in \Lambda$ can be uniquely associated with a $p \times p$ matrix $A_i(\lambda_i)$. Define $\Theta = \Delta \otimes \Lambda$, and let $\epsilon_{n,p} \in \mathbb{R}$ be fixed (the exact value of $\epsilon_{n,p}$ will be chosen later). For each $\theta = (\gamma, \lambda) \in \Theta$ with $\gamma = (\gamma_1, \ldots, \gamma_r) \in \Delta$ and $\lambda = (\lambda_1, \ldots, \lambda_r) \in \Lambda$, we associate $\theta = (\gamma_1, \ldots, \gamma_r, \lambda_1, \ldots, \lambda_r)$ with a volatility matrix $\mathbf{\Gamma}(\theta)$ by

$$(32) \qquad \mathbf{\Gamma}(\theta) = \mathbf{I}_p + \epsilon_{n,p} \sum_{m=1}^{r} \gamma_m A_m(\lambda_m).$$

For simplicity we assume that $\tau > 1$ in the definition of the parameter space $\mathcal{G}_q(\pi_n(p))$ for the minimax problem $\mathcal{Q}_q(\pi_n(p))$; otherwise we replace $\mathbf{I}_p$ in (32) by $C\mathbf{I}_p$ with a small constant $C > 0$. Finally we define $\mathcal{F}_*$ to be a collection of covariance matrices as

$$(33) \qquad \mathcal{F}_* = \left\{ \mathbf{\Gamma}(\theta) : \mathbf{\Gamma}(\theta) = \mathbf{I}_p + \epsilon_{n,p} \sum_{m=1}^{r} \gamma_m A_m(\lambda_m), \theta = (\gamma, \lambda) \in \Theta \right\}.$$

Note that each matrix $\mathbf{\Gamma} \in \mathcal{F}_*$ has value 1 along the main diagonal and contains an $r \times r$ submatrix, say, A, at the upper right corner, A^T at the lower left corner and 0 elsewhere; each row of the submatrix A is either identically 0 (if the corresponding γ value is 0) or has exactly k nonzero elements with value $\epsilon_{n,p}$.

Now we specify the values of $\epsilon_{n,p}$ and k:

$$(34) \qquad \epsilon_{n,p} = \upsilon \left(\frac{\log p}{\sqrt{n}} \right)^{1/2}, \qquad k = \left\lceil \frac{1}{2} \pi_n(p) \epsilon_{n,p}^{-q} \right\rceil - 1,$$

where υ is a fixed small constant that we require

$$(35) \qquad 0 < \upsilon < \left[\min\left\{ \frac{1}{3}, \tau - 1 \right\} \frac{1}{\aleph} \right]^{1/(1-q)}$$

and

$$(36) \qquad 0 < \upsilon^2 < \frac{\beta - 1}{27 c_\kappa \beta},$$

where $c_\kappa = (2\kappa)^{-1}$ satisfies

$$(37) \qquad \sum_{l=1}^{n} a_l^{-2} \le c_\kappa \sqrt{n},$$

since

$$\sum_{l=1}^{n} a_l^{-2} \le \int_0^n \left[1 + 4\kappa^2 n \sin^2\left(\frac{\pi x}{2(n+1)} \right) \right]^{-2} dx \le \frac{n+1}{\pi \kappa \sqrt{n}} \int_0^\infty \left[1 + v^2 \right]^{-2} dv$$

$$= \frac{\sqrt{n} + 1/\sqrt{n}}{4\kappa}.$$

Note that $\epsilon_{n,p}$ and k satisfy $\max_{j \le p} \sum_{i \ne j} |\Gamma_{ij}|^q \le 2k\epsilon_{n,p}^q \le \pi_n(p)$,

$$(38) \qquad 2k\epsilon_{n,p} \le \pi_n(p)\epsilon_{n,p}^{1-q} \le \aleph \upsilon^{1-q} < \min\{\tfrac{1}{3}, \tau - 1\},$$

and consequently every $\mathbf{\Gamma}(\theta)$ is diagonally dominant and positive definite, and $\|\mathbf{\Gamma}(\theta)\|_2 \le \|\mathbf{\Gamma}(\theta)\|_1 \le 2k\epsilon_{n,p} + 1 < \tau$. Thus we have $\mathcal{F}_* \subset \mathcal{G}_q(\pi_n(p))$.

Step 2: *Apply the general lower bound argument.* Let $\mathbf{U}_l$ be independent with

$$\mathbf{U}_l \sim N\big(0, \mathbf{\Gamma}(\theta) + (a_l - 1)\mathbf{I}_p \big),$$

where $l = 1, \ldots, n$, $\theta \in \Theta$, and we denote the joint distribution by P_θ. Applying Lemma 3 in Cai and Zhou (2012) to the parameter space Θ, we have

$$(39) \qquad \inf_{\check{\mathbf{\Gamma}}} \max_{\theta \in \Theta} \mathbb{E}_\theta \|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}(\theta)\|_2^2 \ge \alpha \cdot \frac{r}{8} \cdot \min_{1 \le i \le r} \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\|,$$

where we use $\|\mathbb{P}\|$ to denote the total variation of $\mathbb{P}$,

$$(40) \qquad \alpha \equiv \min_{\{(\theta,\theta') : H(\gamma(\theta), \gamma(\theta')) \ge 1\}} \frac{\|\mathbf{\Gamma}(\theta) - \mathbf{\Gamma}(\theta')\|_2^2}{H(\gamma(\theta), \gamma(\theta'))},$$

$$H\big(\gamma(\theta), \gamma(\theta')\big) = \sum_{i=1}^{r} |\gamma_i(\theta) - \gamma_i(\theta')|$$

and

$$(41) \qquad \bar{\mathbb{P}}_{i,a} = \frac{1}{2^{r-1} D_\Lambda} \sum_{\theta \in \Theta} \mathbb{P}_\theta \cdot \{\theta : \gamma_i(\theta) = a\},$$

where $a \in \{0, 1\}$ and $D_\Lambda = \text{Card}\{\Lambda\}$.

Step 3: *Bound the affinity and per comparison loss.* We need to bound the two factors α and $\min_i \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\|$ in (39). A lower bound for α is given by the following proposition whose proof is the same as that of Lemma 5 in Cai and Zhou (2012).

PROPOSITION 8. *For α defined in equation* (40) *we have*

$$\alpha \geq \frac{(k\epsilon_{n,p})^2}{p}.$$

A lower bound for $\min_i \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\|$ is provided by the proposition below. Since its proof is long and very much involved, the proof details are collected in Appendix II.

PROPOSITION 9. *Let $\mathbf{U}_l$ be independent with $\mathbf{U}_l \sim N(0, \mathbf{\Gamma}(\theta) + (a_l - 1)\mathbf{I}_p)$, $l = 1, \ldots, n$, with $\theta \in \Theta$ and denote the joint distribution by $\mathbb{P}_\theta$. For $a \in \{0, 1\}$ and $1 \leq i \leq r$, define $\bar{\mathbb{P}}_{i,a}$ as in* (41). *Then there exists a constant $C_1 > 0$ such that*

$$\min_{1 \leq i \leq r} \|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\| \geq C_1$$

uniformly over Θ.

Step 4: *Obtain the minimax lower bound.* We obtain the minimax lower bound for estimating $\mathbf{\Gamma}$ over $\mathcal{G}_q(\pi_n(p))$ by combining together (39) and the bounds in Propositions 8 and 9,

$$\inf_{\check{\mathbf{\Gamma}}} \sup_{\mathcal{G}_q(\pi_n(p))} \mathbb{E}\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_2^2 \geq \inf_{\check{\mathbf{\Gamma}}} \max_{\mathbf{\Gamma}(\theta) \in \mathcal{F}_*} \mathbb{E}_\theta \|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}(\theta)\|_2^2 \geq \frac{(k\epsilon_{n,p})^2}{p} \cdot \frac{r}{8} \cdot C_1$$

$$\geq \frac{C_1}{16}(k\epsilon_{n,p})^2 = C_2\pi_n^2(p)\big(n^{-1/4}\sqrt{\log p}\big)^{2-2q}$$

for some constant $C_2 > 0$.

6.3. *Proof of* (16) *for optimal convergence rate under general matrix norm.* The Riesz–Thorin interpolation theorem [Thorin (1948)] implies for $1 \leq d_1 \leq d \leq d_2 \leq \infty$,

$$(42) \qquad \qquad \|\mathbf{A}\|_d \leq \max\{\|\mathbf{A}\|_{d_1}, \|\mathbf{A}\|_{d_2}\}.$$

Set $d_1 = 1$ and $d_2 = \infty$, then (42) yields $\|\mathbf{A}\|_d \leq \max\{\|\mathbf{A}\|_1, \|\mathbf{A}\|_\infty\}$ for $1 \leq d \leq \infty$. When $\mathbf{A}$ is symmetric, (8) shows that $\|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$. Then immediately we have $\|\mathbf{A}\|_d \leq \|\mathbf{A}\|_1$, which means that for a symmetric matrix estimator,

an upper bound under the matrix ℓ_1 norm is also an upper bound under the general matrix ℓ_d norm. Thus, as $\widehat{\boldsymbol{\Gamma}}$ is symmetric, Theorem 2 indicates that for $1 \leq d \leq \infty$,

$$\sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_d^2 \leq C^* \big[\pi_n(p) \big(n^{-1/4} \sqrt{\log p} \big)^{1-q} \big]^2.$$

Now consider the lower bound under the general matrix ℓ_d norm for $1 \leq d \leq \infty$. We will show

$$\inf_{\check{\boldsymbol{\Gamma}}_s} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\check{\boldsymbol{\Gamma}}_s - \boldsymbol{\Gamma}\|_d^2 \geq \inf_{\check{\boldsymbol{\Gamma}}} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_d^2$$

(43)

$$\geq \frac{1}{4} \inf_{\check{\boldsymbol{\Gamma}}_s} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\check{\boldsymbol{\Gamma}}_s - \boldsymbol{\Gamma}\|_d^2,$$

where $\check{\boldsymbol{\Gamma}}$ denotes any matrix estimators of $\boldsymbol{\Gamma}$, and $\check{\boldsymbol{\Gamma}}_s$ any symmetric matrix estimators of $\boldsymbol{\Gamma}$. (43) indicates that it is enough to consider estimators of symmetric matrices.

For symmetric $\mathbf{A}$, (9) shows that $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$. For $d \in (1, \infty)$, $1/d + (d-1)/d = 1$, by duality we have $\|\mathbf{A}\|_d = \|\mathbf{A}\|_{d/(d-1)}$. Also since 2 is always between d and $d/(d-1)$, applying (42) we obtain that $\|\mathbf{A}\|_2 \leq \max\{\|\mathbf{A}\|_d, \|\mathbf{A}\|_{d/(d-1)}\} = \|\mathbf{A}\|_d$. This means that within the class of symmetric matrix estimators, a lower bound under the matrix ℓ_2 norm is also a lower bound under the general matrix ℓ_d norm. Thus (43) and Theorem 3 together imply that for $1 \leq d \leq \infty$,

$$\inf_{\check{\boldsymbol{\Gamma}}} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_d^2 \geq \frac{C_*}{4} \big[\pi_n(p) \big(n^{-1/4} \sqrt{\log p} \big)^{1-q} \big]^2.$$

To complete the proof we need to prove (43). The first inequality of (43) is obvious. For a given matrix estimator $\check{\boldsymbol{\Gamma}}$ we project it onto the parameter space of the minimax problem $\mathcal{P}_q(\pi_n(p))$ by minimizing the matrix ℓ_d norm of $\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_*$ over all $\boldsymbol{\Gamma}_*$ in the parameter space. Denote its projection by $\check{\boldsymbol{\Gamma}}_p$. Since the parameter space consists of symmetric matrices, $\check{\boldsymbol{\Gamma}}_p$ is symmetric. Hence

$$\inf_{\check{\boldsymbol{\Gamma}}_s} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\check{\boldsymbol{\Gamma}}_s - \boldsymbol{\Gamma}\|_d^2$$

$$\leq \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\check{\boldsymbol{\Gamma}}_p - \boldsymbol{\Gamma}\|_d^2$$

$$\leq 2 \sup_{\mathcal{P}_q(\pi_n(p))} \big[\mathbb{E} \|\check{\boldsymbol{\Gamma}}_p - \check{\boldsymbol{\Gamma}}\|_d^2 + \mathbb{E} \|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_d^2 \big]$$

$$\leq 2 \sup_{\mathcal{P}_q(\pi_n(p))} \big[\mathbb{E} \|\boldsymbol{\Gamma} - \check{\boldsymbol{\Gamma}}\|_d^2 + \mathbb{E} \|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_d^2 \big]$$

$$\leq 4 \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_d^2,$$

where the second inequality is from the triangle inequality and the third one follows from the definition of $\check{\Gamma}_p$. Since the above inequality holds for every $\check{\Gamma}$, we have

$$
\inf_{\check{\Gamma}_s} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\|\check{\Gamma}_s - \Gamma\|_d^2
$$

$$
\leq 4 \inf_{\check{\Gamma}} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\|\check{\Gamma} - \Gamma\|_d^2,
$$

which is equivalent to the second inequality of (43).

APPENDIX I: PROOFS OF PROPOSITIONS 5–7

I.1. Proof of Proposition 5. From the expression of V_{ij} in terms of $V_{ij}^{K_m}$ given by (25), we have

$$
P(|V_{ij} - \Gamma_{ij}| \geq d)
$$

$$
\leq P\left(\sum_{m=1}^{N} |a_m||V_{ij}^{K_m} - \Gamma_{ij}| + \zeta(|V_{ij}^{K_1} - \Gamma_{ij}| + |V_{ij}^{K_N} - \Gamma_{ij}|) \geq d\right)
$$

$$
\leq P\left(\sum_{m=1}^{N} |a_m||V_{ij}^{K_m} - \Gamma_{ij}| \geq d/2\right)
$$

(44)
$$
+ P(\zeta|V_{ij}^{K_1} - \Gamma_{ij}| + \zeta|V_{ij}^{K_N} - \Gamma_{ij}| \geq d/2)
$$

$$
\leq \sum_{m=1}^{N} P(|V_{ij}^{K_m} - \Gamma_{ij}| \geq d/(2A)) + P(\zeta|V_{ij}^{K_1} - \Gamma_{ij}| \geq d/4)
$$

$$
+ P(\zeta|V_{ij}^{K_N} - \Gamma_{ij}| \geq d/4),
$$

where $A = \sum_{m=1}^{N} |a_m| = 9/2 + o(1)$.

The definition of $V_{ij}^{K_m}$ in (24) shows

$$
V_{ij}^{K_m} = \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} \{X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})\}\{X_j(\tau_r^{k_m}) - X_j(\tau_{r-1}^{k_m})\}
$$

$$
\equiv \frac{1}{K_m} \sum_{k_m=1}^{K_m} [X_i, X_j]^{(k_m)}
$$

and

$$
V_{ij}^{K_m} - \Gamma_{ij} = \frac{1}{K_m} \sum_{k_m=1}^{K_m} \left[[X_i, X_j]^{(k_m)} - \int_0^1 \gamma_{ij}(s)\, ds\right].
$$

With the above expression for $V_{ij}^{K_m} - \Gamma_{ij}$ we obtain that for $d_1 > 0$ and $1 \le m \le N$,

$$P(|V_{ij}^{K_m} - \Gamma_{ij}| \ge d_1) \le P\left(\frac{1}{K_m} \sum_{k_m=1}^{K_m} \left|[X_i, X_j]^{(k_m)} - \int_0^1 \gamma_{ij}(s)\,ds\right| \ge d_1\right)$$

(45)
$$\le \sum_{k_m=1}^{K_m} P\left(\left|[X_i, X_j]^{(k_m)} - \int_0^1 \gamma_{ij}(s)\,ds\right| \ge d_1\right)$$

$$\le C_1 K_m \exp\left\{-\frac{n}{K_m}\frac{d_1^2}{C_2}\right\} \le C_3 \sqrt{n} \exp\{-\sqrt{n}d_1^2/C_4\},$$

where the third inequality is from Lemma 10 below and the last inequality is due to the fact that $\sqrt{n} \le K_m \le 2\sqrt{n}$ and the maximum distance between consecutive grids in τ^{k_m} is bounded by $K_m/n \le 2/\sqrt{n}$.

Substituting (45) into (44) we immediately prove Proposition 5 as follows:

$$P(|V_{ij} - \Gamma_{ij}| \ge d) \le C_3 N \sqrt{n} \exp\{-\sqrt{n}d^2/(4A^2 C_4)\}$$
$$+ 2C_3 \sqrt{n} \exp\{-\sqrt{n}d^2/(16\zeta^2 C_4)\}$$
$$\le C_5 n \exp\{-\sqrt{n}d^2/C_6\}.$$

LEMMA 10. *Under model* (1) *and condition* A2, *for any sequence* $0 = v_0 \le v_1 < v_2 < \cdots < v_m \le v_{m+1} = 1$ *satisfying* $\max_{1 \le r \le m+1} |v_r - v_{r-1}| \le C/m$, *we have for* $1 \le i, j \le p$ *and small* $d > 0$,

$$P\left(\left|\sum_{r=2}^{m}(X_i(v_r) - X_i(v_{r-1}))(X_j(v_r) - X_j(v_{r-1})) - \int_0^1 \gamma_{ij}(s)\,ds\right| \ge d\right)$$
$$\le C_1 \exp(-md^2/C_2).$$

PROOF. Let $X_i^*(t) = X_i(t) - \int_0^t \mu_{is}\,ds$ and $\mathbf{X}^*(t) = (X_1^*(t), \ldots, X_p^*(t))^T$. Then $\mathbf{X}^*(t)$ is a stochastic integral with respect to $\mathbf{B}_t$ and has the same quadratic variation as $\mathbf{X}(t)$. Let $\mathbf{B}_t = (B_1(t), \ldots, B_p(t))^T$. With $\boldsymbol{\sigma}_t = (\sigma_{ij}(t))$ and $\boldsymbol{\gamma}(t) = (\gamma_{ij}(t)) = \boldsymbol{\sigma}_t^T \boldsymbol{\sigma}_t$ we have

$$X_i^*(t) = \int_0^t \sum_{\ell=1}^p \sigma_{\ell i}(s)\,dB_\ell(s), \qquad i = 1, \ldots, p,$$

with quadratic variation $\langle X_i^*, X_i^* \rangle_t = \int_0^t \gamma_{ii}(s)\,ds$. Also $X_i^* \pm X_j^*$ have quadratic variations

$$\langle X_i^* \pm X_j^*, X_i^* \pm X_j^* \rangle_t = \int_0^1 [\gamma_{ii}(s) + \gamma_{jj}(s) \pm 2\gamma_{ij}(s)]\,ds.$$

Define

$$B_i^*(t) = \int_0^t \gamma_{ii}^{-1/2}(s) \sum_{\ell=1}^p \sigma_{\ell i}(s) \, dB_\ell(s).$$

Then

$$X_i^*(t) = \int_0^t \gamma_{ii}^{1/2}(s) \, dB_i^*(s),$$

B_i^* is a continuous-time martingale and has quadratic variation

$$\langle B_i^*, B_i^* \rangle_t = \int_0^t \gamma_{ii}^{-1}(s) \sum_{\ell=1}^p \sigma_{\ell i}^2(s) \, ds = \int_0^t \gamma_{ii}^{-1}(s)\gamma_{ii}(s) \, ds = t,$$

and hence Lévy's martingale characterization of Brownian motion shows that B_i^* is a one-dimensional Brownian motion; see Karatzas and Shreve [(1991), Theorem 3.16]. We can apply Lemma 3 in Fan, Li and Yu (2012) to each X_i^* and obtain for $1 \le i \le p$,

(46)
$$P\left(\left| \sum_{r=2}^m [X_i^*(v_r) - X_i^*(v_{r-1})]^2 - \int_{v_1}^{v_m} \gamma_{ii}(s) \, ds \right| \ge d \right)$$
$$\le 4 \exp\{-md^2/C_0\}.$$

Similarly for $X_i^* \pm X_j^*$, we define

$$B_{ij}^\pm(s) = \int_0^t [\gamma_{ii}(s) + \gamma_{jj}(s) \pm 2\gamma_{ij}(s)]^{-1/2} \sum_{\ell=1}^p [\sigma_{\ell i}(s) \pm \sigma_{\ell j}(s)] \, dB_\ell(s).$$

Then

$$X_i^*(t) \pm X_j^*(t) = \int_0^t [\gamma_{ii}(s) + \gamma_{jj}(s) \pm 2\gamma_{ij}(s)]^{1/2} \, dB_{ij}^\pm(s),$$

$B_{ij}^\pm$ are continuous-time martingales with quadratic variations

$$\langle B_{ij}^\pm, B_{ij}^\pm \rangle_t = \int_0^t [\gamma_{ii}(s) + \gamma_{jj}(s) \pm 2\gamma_{ij}(s)]^{-1}$$
$$\times \sum_{\ell=1}^p [\sigma_{\ell i}^2(s) + \sigma_{\ell j}^2(s) \pm 2\sigma_{\ell i}(s)\sigma_{\ell j}(s)] \, ds$$
$$= \int_0^t [\gamma_{ii}(s) + \gamma_{jj}(s) \pm 2\gamma_{ij}(s)]^{-1} [\gamma_{ii}(s) + \gamma_{jj}(s) \pm 2\gamma_{ij}(s)] \, ds = t,$$

and hence Lévy's martingale characterization of Brownian motion implies that $B_{ij}^\pm$ are one-dimensional Brownian motions. We can apply Lemma 3 in Fan, Li and Yu

(2012) to each of $X_i^* + X_j^*$ and $X_i^* - X_j^*$ and obtain for $1 \le i, j \le p$,

$$P\left(\left|\sum_{r=2}^m ([X_i^*(\nu_r) - X_i^*(\nu_{r-1})] \pm [X_j^*(\nu_r) - X_j^*(\nu_{r-1})])^2\right.\right.$$

(47)
$$\left.\left. - \int_{\nu_1}^{\nu_m} [\gamma_{ii}(s) + \gamma_{jj}(s) \pm 2\gamma_{ij}(s)]\,ds\right| \ge d\right)$$

$$\le 4\exp\{-md^2/C_0\}.$$

Note that

$$4\gamma_{ij}(s) = [\gamma_{ii}(s) + \gamma_{jj}(s) + 2\gamma_{ij}(s)] - [\gamma_{ii}(s) + \gamma_{jj}(s) - 2\gamma_{ij}(s)],$$

$$4\sum_{r=2}^m (X_i^*(\nu_r) - X_i^*(\nu_{r-1}))(X_j^*(\nu_r) - X_j^*(\nu_{r-1}))$$

$$= \sum_{r=2}^m \{[X_i^*(\nu_r) - X_i^*(\nu_{r-1})] + [X_j^*(\nu_r) - X_j^*(\nu_{r-1})]\}^2$$

$$- \sum_{r=2}^m \{[X_i^*(\nu_r) - X_i^*(\nu_{r-1})] - [X_j^*(\nu_r) - X_j^*(\nu_{r-1})]\}^2,$$

and thus

$$4\left|\sum_{r=2}^m (X_i^*(\nu_r) - X_i^*(\nu_{r-1}))(X_j^*(\nu_r) - X_j^*(\nu_{r-1})) - \int_{\nu_1}^{\nu_m} \gamma_{ij}(s)\,ds\right|$$

$$\le \left|\sum_{r=2}^m \{[X_i^*(\nu_r) - X_i^*(\nu_{r-1})] + [X_j^*(\nu_r) - X_j^*(\nu_{r-1})]\}^2\right.$$

$$\left. - \int_{\nu_1}^{\nu_m} [\gamma_{ii}(s) + \gamma_{jj}(s) + 2\gamma_{ij}(s)]\,ds\right|$$

$$+ \left|\sum_{r=2}^m \{[X_i^*(\nu_r) - X_i^*(\nu_{r-1})] - [X_j^*(\nu_r) - X_j^*(\nu_{r-1})]\}^2\right.$$

$$\left. - \int_{\nu_1}^{\nu_m} [\gamma_{ii}(s) + \gamma_{jj}(s) - 2\gamma_{ij}(s)]\,ds\right|.$$

Combining (47) and above inequality we conclude

$$P\left(\left|\sum_{r=2}^m (X_i^*(\nu_r) - X_i^*(\nu_{r-1}))(X_j^*(\nu_r) - X_j^*(\nu_{r-1})) - \int_{\nu_1}^{\nu_m} \gamma_{ij}(s)\,ds\right| \ge d\right)$$

(48)
$$\le 8\exp\{-m(d/8)^2/C_0\} = 8\exp\{-md^2/(64C_0)\}.$$

On the other hand,

$$\sum_{r=2}^{m}(X_i(v_r) - X_i(v_{r-1}))(X_j(v_r) - X_j(v_{r-1}))$$

$$= \sum_{r=2}^{m}\left\{[X_i^*(v_r) - X_i^*(v_{r-1})] + \int_{v_{r-1}}^{v_r}\mu_{is}\,ds\right\}$$

$$\times\left\{[X_j^*(v_r) - X_j^*(v_{r-1})] + \int_{v_{r-1}}^{v_r}\mu_{js}\,ds\right\}$$

$$(49) \qquad = \sum_{r=2}^{m}(X_i^*(v_r) - X_i^*(v_{r-1}))(X_j^*(v_r) - X_j^*(v_{r-1}))$$

$$+ \sum_{r=2}^{m}\int_{v_{r-1}}^{v_r}\mu_{is}\,ds\int_{v_{r-1}}^{v_r}\mu_{js}\,ds$$

$$+ \sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]\int_{v_{r-1}}^{v_r}\mu_{js}\,ds$$

$$+ \sum_{r=2}^{m}[X_j^*(v_r) - X_j^*(v_{r-1})]\int_{v_{r-1}}^{v_r}\mu_{is}\,ds.$$

From condition A2 we have that μ_i and μ_j are bounded by c_1, and thus

$$(50) \qquad \left|\sum_{r=2}^{m}\int_{v_{r-1}}^{v_r}\mu_{is}\,ds\int_{v_{r-1}}^{v_r}\mu_{js}\,ds\right| \le \frac{c_1^2}{m}.$$

Applications of Hölder's inequality lead to

$$\left|\sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]\int_{v_{r-1}}^{v_r}\mu_{js}\,ds\right|^2$$

$$(51) \qquad \le \sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]^2\sum_{r=2}^{m}\left|\int_{v_{r-1}}^{v_r}\mu_{js}\,ds\right|^2$$

$$\le \frac{c_1^2}{m}\sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]^2,$$

$$\left|\sum_{r=2}^{m}[X_j^*(v_r) - X_j^*(v_{r-1})]\int_{v_{r-1}}^{v_r}\mu_{is}\,ds\right|^2$$

$$(52) \qquad \le \frac{c_1^2}{m}\sum_{r=2}^{m}[X_j^*(v_r) - X_j^*(v_{r-1})]^2.$$

From (49) we have

$$
P\left(\left|\sum_{r=2}^{m}(X_i(v_r) - X_i(v_{r-1}))(X_j(v_r) - X_j(v_{r-1})) - \int_{v_1}^{v_m}\gamma_{ij}(s)\,ds\right| \geq d\right)
$$

$$
\leq P\left(\left|\sum_{r=2}^{m}(X_i^*(v_r) - X_i^*(v_{r-1}))(X_j^*(v_r) - X_j^*(v_{r-1}))\right.\right.
$$

$$
\left.\left. - \int_{v_1}^{v_m}\gamma_{ij}(s)\,ds\right| \geq d/4\right)
$$

$$
+ P\left(\left|\sum_{r=2}^{m}\int_{v_{r-1}}^{v_r}\mu_{is}\,ds\int_{v_{r-1}}^{v_r}\mu_{js}\,ds\right| \geq d/4\right)
$$

(53)
$$
+ P\left(\left|\sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]\int_{v_{r-1}}^{v_r}\mu_{js}\,ds\right| \geq d/4\right)
$$

$$
+ P\left(\left|\sum_{r=2}^{m}[X_j^*(v_r) - X_j^*(v_{r-1})]\int_{v_{r-1}}^{v_r}\mu_{is}\,ds\right| \geq d/4\right)
$$

$$
\leq 8\exp\{-m(d/4)^2/(64C_0)\} + 1\left(\frac{c_1^2}{m} \geq d/4\right)
$$

$$
+ P\left(\sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]^2 \geq md^2/(16c_1^2)\right)
$$

$$
+ P\left(\sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]^2 \geq md^2/(16c_1^2)\right),
$$

where the last inequality is due to the bounds obtained from (48) and (50)–(52) for the four respective probability terms. We handle the last two terms on the right-hand side of (53) as follows. If $md^2/(16c_1^2) - c_2 > 0$ [or equivalently $d > 4c_1(c_2/m)^{1/2}$], using condition A2 (which implies $\gamma_{ii} \leq c_2$ and $\gamma_{jj} \leq c_2$) and (46), we get

$$
P\left(\sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]^2 \geq md^2/(16c_1^2)\right)
$$

$$
+ P\left(\sum_{r=2}^{m}[X_j^*(v_r) - X_j^*(v_{r-1})]^2 \geq md^2/(16c_1^2)\right)
$$

(54)
$$
\leq P\left(\sum_{r=2}^{m}[X_i^*(v_r) - X_i^*(v_{r-1})]^2 - \int_{v_1}^{v_m}\gamma_{ii}(s)\,ds \geq md^2/(16c_1^2) - c_2\right)
$$

$$+ P\left(\sum_{r=2}^{m}[X_j^*(\nu_r) - X_j^*(\nu_{r-1})]^2 - \int_{\nu_1}^{\nu_m} \gamma_{jj}(s)\,ds \geq md^2/(16c_1^2) - c_2\right)$$

$$\leq P\left(\left|\sum_{r=2}^{m}[X_i^*(\nu_r) - X_i^*(\nu_{r-1})]^2 - \int_{\nu_1}^{\nu_m} \gamma_{ii}(s)\,ds\right| \geq md^2/(16c_1^2) - c_2\right)$$

$$+ P\left(\left|\sum_{r=2}^{m}[X_j^*(\nu_r) - X_j^*(\nu_{r-1})]^2 - \int_{\nu_1}^{\nu_m} \gamma_{jj}(s)\,ds\right| \geq md^2/(16c_1^2) - c_2\right)$$

$$\leq 8\exp\{-m[md^2/(16c_1^2) - c_2]^2/C_0\},$$

which is bounded by $8\exp\{-md^2/C_0\}$, if $m[md^2/(16c_1^2) - c_2]^2 > md^2$, which is true provided that

$$(55) \qquad\qquad d > \frac{8c_1^2}{m} + \frac{4c_1}{m}(4c_1^2 + mc_2)^{1/2}.$$

Putting together (53) and the probability bound from (54)–(55), we conclude that if

$$d > \max\left\{\frac{4c_1^2}{m}, \frac{4c_1 c_2^{1/2}}{m^{1/2}}, \frac{8c_1^2}{m} + \frac{4c_1}{m}(4c_1^2 + mc_2)^{1/2}\right\}$$

$$= \frac{8c_1^2}{m} + \frac{4c_1}{m}(4c_1^2 + mc_2)^{1/2},$$

$$P\left(\left|\sum_{r=2}^{m}(X_i(\nu_r) - X_i(\nu_{r-1}))(X_j(\nu_r) - X_j(\nu_{r-1}))\right.\right.$$

$$\left.\left. - \int_{\nu_1}^{\nu_m} \gamma_{ij}(s)\,ds\right| \geq d\right)$$

$$(56)$$

$$\leq 8\exp\{-md^2/(1024C_0)\} + 8\exp\{-md^2/C_0\}$$

$$\leq 16\exp\{-md^2/(1024C_0)\}.$$

From condition A2 we have $|\gamma_{ij}| \leq (\gamma_{ii}\gamma_{jj})^{1/2} \leq c_2$ and

$$\left|\int_{\nu_1}^{\nu_m} \gamma_{ij}(s)\,ds - \int_0^1 \gamma_{ij}(s)\,ds\right|$$

$$\leq c_2(\nu_1 + 1 - \nu_m) \leq 2c_2/m.$$

Then (56) and above inequality imply that if

$$(57) \qquad\qquad d > \max\left\{\frac{4c_2}{m}, \frac{8c_1^2}{m} + \frac{4c_1}{m}(4c_1^2 + mc_2)^{1/2}\right\},$$

$$P\left(\left|\sum_{r=2}^{m}(X_i(v_r) - X_i(v_{r-1}))(X_j(v_r) - X_j(v_{r-1})) - \int_0^1 \gamma_{ij}(s)\,ds\right| \geq d\right)$$

$$\leq P\left(\left|\sum_{r=2}^{m}(X_i(v_r) - X_i(v_{r-1}))(X_j(v_r) - X_j(v_{r-1}))\right.\right.$$

$$\left.\left. - \int_{v_1}^{v_m} \gamma_{ij}(s)\,ds\right| \geq d/2\right)$$

$$\leq 16\exp\{-m(d/2)^2/(1024C_0)\} = 16\exp\{-md^2/(4096C_0)\}.$$

This proves the lemma with $C_1 = 16$ and $C_2 = 4096C_0$ for d satisfies (57).

If (57) is not satisfied, we have

$$d \leq \max\left\{\frac{4c_2}{m}, \frac{8c_1^2}{m} + \frac{4c_1}{m}(4c_1^2 + mc_2)^{1/2}\right\} \leq \frac{8c_1^2 + 4c_2 + 4c_1c_2^{1/2}}{m^{1/2}} \equiv \frac{C}{m^{1/2}}.$$

Then the tail probability bound in the lemma obeys

$$C_1\exp\{-md^2/C_2\} \geq C_1\exp\{-C^2/C_2\},$$

and we easily show the probability inequality in the lemma by choosing $C_1 = C_1'$ and $C_2 = C_2'$, where C_1' and C_2' satisfy $C_1'\exp\{-C^2/C_2'\} \geq 1$.

Finally taking $C_1 = \max(16, C_1')$ and $C_2 = \max(4096C_0, C_2')$ we establish the tail probability, regardless whether d satisfies (57) or not, and complete the proof. $\square$

I.2. Proof of Proposition 6.   As the proofs for $G_{ij}(2)$ and $G_{ij}(3)$ are similar, we give arguments only for $G_{ij}(2)$. Lemma 11 below establishes the tail probability for $G_{ij}^{K_m}(2)$. Using the expression of $G_{ij}(2)$ in terms of $G_{ij}^{K_m}(2)$ given by (25) and applying Lemma 11, we obtain

$$P(|G_{ij}(2)| \geq d)$$

$$\leq \sum_{m=1}^{N} P(|G_{ij}^{K_m}(2)| \geq d/(2A)) + P(\zeta|G_{ij}^{K_1}(2)| \geq d/4)$$

$$+ P(\zeta|G_{ij}^{K_N}(2)| \geq d/4)$$

$$\leq C_1 N\sqrt{n}\exp\{-\sqrt{n}d^2/(4A^2C_2)\} + 2C_1\sqrt{n}\exp\{-\sqrt{n}d^2/(16\zeta^2C_2)\}$$

$$\leq C_3 n\exp\{-\sqrt{n}d^2/C_4\},$$

where $A = \sum_{m=1}^{N}|a_m| = 9/2 + o(1)$.

LEMMA 11.   *Under the assumptions of Theorem 1, we have for $1 \leq i, j \leq p$ and $1 \leq m \leq N$,*

$$P(|G_{ij}^{K_m}(2)| \geq d) \leq C_1\sqrt{n}\exp\{-\sqrt{n}d^2/C_2\}.$$

PROOF.   Simple algebraic manipulations show

$$G_{ij}^{K_m}(2) = \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} [X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})][\varepsilon_j(\tau_r^{k_m}) - \varepsilon_j(\tau_{r-1}^{k_m})]$$

$$= \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} [X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})]\varepsilon_j(\tau_r^{k_m})$$

$$- \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} [X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})]\varepsilon_j(\tau_{r-1}^{k_m})$$

$$\equiv R_5^{K_m} - R_6^{K_m}.$$

The lemma is proved if we establish tail probabilities for both $R_5^{K_m}$ and $R_6^{K_m}$. Due to similarity, we give the arguments only for $R_5^{K_m}$. Since $\mathbf{X}_t$ and $\varepsilon_i(t_\ell)$ are independent, conditional on the whole path of $\mathbf{X}_t$, $R_5^{K_m}$ is the weighted sum of $\varepsilon_j(\cdot)$. Hence,

$$P(|R_5^{K_m}| \geq d)$$

$$= E[P(|R_5^{K_m}| \geq d|\mathbf{X}_t, t \in [0, 1))]$$

$$= E\left[P\left(\left|\sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} [X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})]\varepsilon_j(\tau_r^{k_m})\right| \geq dK_m|\mathbf{X}_t, t \in [0, 1]\right)\right]$$

(58)

$$\leq E\left[c_0 \exp\left\{-\frac{d^2 K_m}{2\tau_0 V_{ii}^{K_m} \eta_j}\right\}\right]$$

$$= E\left[c_0 \exp\left\{-\frac{d^2 K_m}{2\tau_0 V_{ii}^{K_m} \eta_j}\right\}1(\Omega_0)\right] + E\left[c_0 \exp\left\{-\frac{d^2 K_m}{2\tau_0 V_{ii}^{K_m} \eta_j}\right\}1(\Omega_0^c)\right]$$

$$\equiv R_{5,1}^{K_m} + R_{5,2}^{K_m},$$

where the inequality is due to the subGaussianity of $\varepsilon_j(\cdot)$ defined in (10), η_j is the variance of $\varepsilon_j(\cdot)$, $V_{ii}^{K_m}$ is given by (24) with an expression

$$V_{ii}^{K_m} = \frac{1}{K_m} \sum_{k_m=1}^{K_m} [X_i, X_i]^{k_m} = \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} [X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})]^2$$

and

$$\Omega_0 = \{|V_{ii}^{K_m} - \Gamma_{ii}| \geq d\}.$$

From the definition of Ω_0 and conditions A1–A2, we have $\eta_j \leq \kappa^2$, $\Gamma_{ii} \leq c_2$ and $V_{ii}^{K_m} \leq \Gamma_{ii} + d \leq c_2 + d$ on Ω_0^c. Thus for small d we have

$$
\begin{aligned}
R_{5,2}^{K_m} &= E\left[c_0 \exp\left\{-\frac{K_m d^2}{2\tau_0 V_{ii}^{K_m} \eta_j}\right\} 1(\Omega_0^c)\right] \\
&\leq C_1 \exp\{-K_m d^2/C_2\} \leq C_1 \exp\{-\sqrt{n} d^2/C_2\}.
\end{aligned}
$$

(59)

On the other hand, from (45) (in the proof of Proposition 5) we have

$$
P(\Omega_0) \leq C_3 \sqrt{n} \exp\{-\sqrt{n} d^2/C_4\},
$$

and thus

$$
\begin{aligned}
R_{5,1}^{K_m} &= E\left[c_0 \exp\left\{-\frac{d^2 K_m}{2\tau_0 V_{ii}^{K_m} \eta_j}\right\} 1(\Omega_0)\right] \leq c_0 P(\Omega_0) \\
&\leq c_0 C_3 \sqrt{n} \exp\{-\sqrt{n} d^2/C_4\}.
\end{aligned}
$$

(60)

Finally substituting (59) and (60) into (58) we obtain

$$
\begin{aligned}
P(|R_5^{K_m}| \geq d) &\leq C_1 \exp\{-\sqrt{n} d^2/C_2\} + c_0 C_3 \sqrt{n} \exp\{-\sqrt{n} d^2/C_4\} \\
&\leq C_5 \sqrt{n} \exp\{-\sqrt{n} d^2/C_6\}. \qquad \square
\end{aligned}
$$

I.3. Proof of Proposition 7. Denote by $\rho_{ij}(0)$ the correlation between $\varepsilon_i(t_1)$ and $\varepsilon_j(t_1)$. From the expression of $G_{ij}(1)$ in terms of $G_{ij}^{K_m}(1)$ given by (25) we obtain that $P(|G_{ij}(1)| \geq d)$ is bounded by

$$
\begin{aligned}
&P\left(\left|\sum_{m=1}^{N} a_m G_{ij}^{K_m}(1) + 2\sqrt{\eta_i \eta_j} \rho_{ij}(0)\right| \geq d/2\right) \\
&\quad + P\left(\left|\zeta\left(G_{ij}^{K_1}(1) - G_{ij}^{K_N}(1)\right) - 2\sqrt{\eta_i \eta_j} \rho_{ij}(0)\right| \geq d/2\right) \\
&\quad \leq C_1 \sqrt{n} \exp\{-\sqrt{n} d^2/(4C_2)\} + C_3 \exp\{-n d^2/(4C_4)\} \\
&\quad \leq C_5 \sqrt{n} \exp\{-\sqrt{n} d^2/C_6\},
\end{aligned}
$$

where the first inequality is from Lemmas 12 and 13 below.

LEMMA 12.   *Under the assumptions of Theorem 1, we have for $1 \leq i, j \leq p$,*

$$
P\left(\left|\sum_{m=1}^{N} a_m G_{ij}^{K_m}(1) + 2\sqrt{\eta_i \eta_j} \rho_{ij}(0)\right| \geq d\right) \leq C_1 \sqrt{n} \exp\{-\sqrt{n} d^2/C_2\}.
$$

PROOF. From the definition of $\mathbf{G}^{K_m} = (G_{ij}^{K_m}(1))$ in (24), we have

$$G_{ij}^{K_m}(1) = \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} \left[\varepsilon_i(\tau_r^{k_m}) - \varepsilon_i(\tau_{r-1}^{k_m})\right]\left[\varepsilon_j(\tau_r^{k_m}) - \varepsilon_j(\tau_{r-1}^{k_m})\right]$$

$$= \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\tau^{k_m}|} \left[\varepsilon_i(\tau_r^{k_m})\varepsilon_j(\tau_r^{k_m}) - \varepsilon_i(\tau_r^{k_m})\varepsilon_j(\tau_{r-1}^{k_m}) - \varepsilon_i(\tau_{r-1}^{k_m})\varepsilon_j(\tau_r^{k_m})\right.$$

$$\left. + \varepsilon_i(\tau_{r-1}^{k_m})\varepsilon_j(\tau_{r-1}^{k_m})\right]$$

$$= \frac{2}{K_m} \sum_{r=1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_r) - \frac{1}{K_m} \sum_{r=1}^{K_m} \varepsilon_i(t_r)\varepsilon_j(t_r) - \frac{1}{K_m} \sum_{r=n-K_m+1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_r)$$

$$- \frac{1}{K_m} \sum_{r=K_m+1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_{r-K_m}) - \frac{1}{K_m} \sum_{r=K_m+1}^{n} \varepsilon_i(t_{r-K_m})\varepsilon_j(t_r)$$

$$\equiv I_0^{K_m} - I_1^{K_m} - I_2^{K_m} - I_3^{K_m} - I_4^{K_m}$$

and

$$(61) \quad \sum_{m=1}^{N} a_m G_{ij}^{K_m}(1) = \sum_{m=1}^{N} a_m I_0^{K_m} - \sum_{i=1}^{4} \sum_{m=1}^{N} a_m I_i^{K_m} \equiv I_0 - I_1 - I_2 - I_3 - I_4.$$

Note that $\sum_{m=1}^{N} a_m/K_m = 0$, and

$$I_0 = \sum_{m=1}^{N} a_m I_0^{K_m} = \sum_{m=1}^{N} \frac{a_m}{K_m} \sum_{r=1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_r) = 0.$$

Hence,

$$(62) \quad \begin{aligned} &P\left(\left|\sum_{m=1}^{N} a_m G_{ij}^{K_m}(1) + 2\sqrt{\eta_i \eta_j}\rho_{ij}(0)\right| \geq d\right) \\ &\leq \sum_{i=1}^{2} P\left(|I_i - \sqrt{\eta_i \eta_j}\rho_{ij}(0)| \geq d/4\right) + \sum_{i=3}^{4} P\left(|I_i| \geq d/4\right). \end{aligned}$$

To prove the lemma we need to derive the four tail probabilities on the right-hand side of (62). Below we will establish the tail probabilities for I_1, I_2, I_3 and I_4 by using large deviation results for the case of m-dependent random variables in Saulis and Statulevičius (1991). Because of similarity, we give arguments only for the tail probabilities of I_1 and I_3.

First for I_1, from the definition of a_m in (6) we have

$$I_1 - \sqrt{\eta_i \eta_j}\rho_{ij}(0) = \sum_{m=1}^{N} a_m \left[I_1^{K_m} - \sqrt{\eta_i \eta_j}\rho_{ij}(0)\right],$$

$$(63) \quad P\big(|I_1 - \sqrt{\eta_i \eta_j} \rho_{ij}(0)| \geq d/4\big) \leq \sum_{m=1}^{N} P\big(|I_1^{K_m} - \sqrt{\eta_i \eta_j} \rho_{ij}(0)| \geq d/(4A)\big),$$

where $A = \sum_{m=1}^{N} |a_m| = 9/2 + o(1)$. The M-dependence of $(\varepsilon_1(t_\ell), \ldots, \varepsilon_p(t_\ell))$ in condition A1 indicates that $\varepsilon_i(t_r)\varepsilon_j(t_r)$, $r = 1, \ldots, n$, are M-dependent, $I_1^{K_m}$ is the average of $\varepsilon_i(t_r)\varepsilon_j(t_r)$, $r = 1, \ldots, K_m$, and Lemma 14 below calculates $E(I_1^{K_m}) = \sqrt{\eta_i \eta_j} \rho_{ij}(0)$ and $\mathrm{Var}(I_1^{K_m}) \leq Cn^{-1/2}$. Also for any integer k,

$$
\begin{aligned}
E\big(|\varepsilon_i(t_r)\varepsilon_j(t_r)|^k\big) &\leq \sqrt{E\big(|\varepsilon_i(t_r)|^{2k}\big) E\big(|\varepsilon_j(t_r)|^{2k}\big)} \\
&\leq c_0(2k)!(2\tau_0)^{2k} \leq c_0(k!)^2(16\tau_0^2)^k \leq (k!)^2\big[16\tau_0^2(c_0 \vee 1)\big]^k,
\end{aligned}
$$

where the first inequality is from the Cauchy–Schwarz inequality, and the second inequality is from the subGaussian tails of $\varepsilon_i(t_r)$ and $\varepsilon_j(t_r)$, which imply that their $2k$-moments are bounded by $\int_0^\infty c_0 \exp[-x^{1/(2k)}/(2\tau_0)]\,dx = c_0(2k)!(2\tau_0)^{2k}$. Applying Theorem 4.30 in Saulis and Statulevičius (1991) to M-dependent random variables $\varepsilon_i(t_r)\varepsilon_j(t_r)$ we obtain

$$(64) \qquad P\big(|I_1^{K_m} - \sqrt{\eta_i \eta_j} \rho_{ij}(0)| \geq d_1\big) \leq C_1 \exp\left\{-\frac{\sqrt{n}d_1^2}{C_2}\right\}.$$

Plugging (64) with $d_1 = d/(4A)$ into (63) we establish the tail probability for I_1

$$
\begin{aligned}
P\big(|I_1 - \sqrt{\eta_i \eta_j} \rho_{ij}(0)| \geq d/4\big) &\leq C_1 N \exp\left\{-\frac{\sqrt{n}d^2}{16A^2 C_2}\right\} \\
&\leq C_3 \sqrt{n} \exp\left\{-\frac{\sqrt{n}d^2}{C_4}\right\}.
\end{aligned}
$$

(65)

Second, consider I_3. We may express it as follows:

$$I_3 = \sum_{m=1}^{N} \sum_{r=K_m+1}^{n} \frac{a_m}{K_m} \varepsilon_i(t_r)\varepsilon_j(t_{r-K_m}) = \sum_{r=1}^{n-K_1} \sum_{m=1}^{(n-N-r)\wedge N} \frac{a_m}{K_m} \varepsilon_j(t_r)\varepsilon_i(t_{r+K_m}),$$

and Lemma 14 below derives $E(I_3) = 0$ and $\mathrm{Var}(I_3) \leq Cn^{-1/2}$.

As $(\varepsilon_1(t_\ell), \ldots, \varepsilon_p(t_\ell))$, $\ell = 1, \ldots, n$, are serially M-dependent, that is, for any integers k and k', and integer sets $\{\ell_1, \ldots, \ell_k\}$ and $\{\ell'_1, \ldots, \ell'_{k'}\}$, $\{\varepsilon_i(t_{\ell_1}), \ldots, \varepsilon_i(t_{\ell_k}), i = 1, \ldots, p\}$ and $\{\varepsilon_i(t_{\ell'_1}), \ldots, \varepsilon_i(t_{\ell'_{k'}}), i = 1, \ldots, p\}$ are independent if every integer in $\{\ell_1, \ldots, \ell_k\}$ differs by more than M from any integer in $\{\ell'_1, \ldots, \ell'_{k'}\}$. Since $K_m > M$ for n large enough, if integers r and r' differ by more than $K_N + M$, for two integer sets $\{r, r+K_m; m = 1, \ldots, (n-N-r)\wedge N\}$ and $\{r', r'+K_m; m = 1, \ldots, (n-N-r')\wedge N\}$, every element in one integer set must be more than M apart from any element in the other integer set. Then $\{\varepsilon_j(t_r), \varepsilon_i(t_{r+K_m}); m = 1, \ldots, (n-N-r)\wedge N\}$ and $\{\varepsilon_j(t_{r'}), \varepsilon_i(t_{r'+K_m}); m = 1, \ldots, (n-N-r')\wedge N\}$ are

independent, and thus $\varepsilon_j(t_r)\varepsilon_i(t_{r+K_m})$, $r = 1, \ldots, n - K_m$, are serially $(K_N + M)$-dependent. Also for any integer k,

$$
E\big(\big|\varepsilon_j(t_r)\varepsilon_i(t_{r+K_m})\big|^k\big) \leq \sqrt{E\big(\big|\varepsilon_j(t_r)\big|^{2k}\big)E\big(\big|\varepsilon_i(t_{r+K_m})\big|^{2k}\big)}
$$
$$
\leq c_0(2k)!(2\tau_0)^{2k} \leq c_0(k!)^2(16\tau_0^2)^k \leq (k!)^2\big[16\tau_0^2(c_0 \vee 1)\big]^k,
$$

where the first inequality is from the Cauchy–Schwarz inequality, and the second inequality is from the subGaussian tails of $\varepsilon_j(t_r)$ and $\varepsilon_i(t_{r+K_m})$. Applying theorem 4.16 in Saulis and Statulevičius (1991) we derive a bound $(k!)^3 C_0^k$ on the kth cumulant of $n^{1/4} I_3$, and then using Lemmas 2.3 and 2.4 in Saulis and Statulevičius (1991) we establish the tail probability for I_3 as follows:

$$
(66) \qquad P\big(|I_3| \geq d/4\big) \leq C_1 \exp\bigg\{-\frac{\sqrt{n}(d/4)^2}{C_2}\bigg\} \leq C_3 \exp\bigg\{-\frac{\sqrt{n}d^2}{C_4}\bigg\}.
$$

Since I_2 and I_4 have the same tail probabilities as I_1 and I_3 given by (65) and (66), respectively, combining them with (62) we conclude

$$
P\bigg(\bigg|\sum_{m=1}^N a_m G_{ij}^{K_m}(1) + 2\sqrt{\eta_i\eta_j}\rho_{ij}(0)\bigg| \geq d\bigg)
$$
$$
\leq 2C_3\sqrt{n}\exp\bigg\{-\frac{\sqrt{n}d^2}{C_4}\bigg\} + 2C_3\exp\bigg\{-\frac{\sqrt{n}d^2}{C_4}\bigg\}
$$
$$
\leq C_5\sqrt{n}\exp\bigg\{-\frac{\sqrt{n}d^2}{C_6}\bigg\}. \qquad \qquad \square
$$

LEMMA 13.   *Under the assumptions of Theorem* 1, *we have for* $1 \leq i, j \leq p$,

$$
(67) \quad P\big(\big|\zeta\big(G_{ij}^{K_1}(1) - G_{ij}^{K_N}(1)\big) - 2\sqrt{\eta_i\eta_j}\rho_{ij}(0)\big| \geq d\big) \leq C_1 \exp\{-nd^2/C_2\}.
$$

PROOF.   First consider $\zeta G_{ij}^{K_1}(1)$ term:

$$
\zeta G_{ij}^{K_1}(1) = \frac{K_N}{n(N-1)} \sum_{k_1=1}^{K_1} \sum_{r=2}^{|\tau^{k_1}|} \big(\varepsilon_i(\tau_r^{k_1}) - \varepsilon_i(\tau_{r-1}^{k_1})\big)\big(\varepsilon_j(\tau_r^{k_1}) - \varepsilon_j(\tau_{r-1}^{k_1})\big)
$$
$$
= \frac{K_N}{n(N-1)} \sum_{r=K_1+1}^{n} \big(\varepsilon_i(t_r)\varepsilon_j(t_r) + \varepsilon_i(t_{r-K_1})\varepsilon_j(t_{r-K_1})
$$
$$
- \varepsilon_i(t_r)\varepsilon_j(t_{r-K_1}) - \varepsilon_i(t_{r-K_1})\varepsilon_j(t_r)\big)
$$
$$
\equiv R_1 + R_2 - R_3 - R_4.
$$

Due to similarity, we show the tail probabilities only for R_1 and R_3. Lemma 14 below calculates the mean and variances of R_1 and R_3. Since R_1 and R_3 have,

respectively, the same structures as $I_1^{K_m}$ and I_3 used in the proof Lemma 12, the arguments for establishing the tail probabilities for $I_1^{K_m}$ and I_3 can be used to derive the tail probability bounds for R_1 and R_3. Consequently we obtain that

$$(68) \quad P\left(\left|\zeta G_{ij}^{K_1}(1) - \frac{2K_N(n-K_1)}{n(N-1)}\sqrt{\eta_i\eta_j}\rho_{ij}(0)\right| \geq d\right) \leq C_1 \exp\{-nd^2/C_2\}.$$

As $G_{ij}^{K_N}(1)$ has the same structure as $\zeta G_{ij}^{K_1}(1)$, similarly we can establish a tail probability for $\zeta G_{ij}^{K_N}(1)$ as follows:

$$(69) \quad P\left(\left|\zeta G_{ij}^{K_N}(1) - \frac{2K_1(n-K_N)}{n(N-1)}\sqrt{\eta_i\eta_j}\rho_{ij}(0)\right| \geq d\right) \leq C_1 \exp\{-nd^2/C_2\}.$$

Since

$$\frac{K_N(n-K_1)}{n(N-1)} - \frac{K_1(n-K_N)}{n(N-1)} = 1,$$

combining (68) and (69) we prove the lemma. $\quad\square$

LEMMA 14. *Under the assumptions of Theorem* 1 *and for large enough n so that $M < K_1$, we have*

$$E(I_3) = E(R_3) = 0, \qquad E\left(I_1^{K_m}\right) = \sqrt{\eta_i\eta_j}\rho_{ij}(0),$$

$$E(R_1) = \frac{K_N(n-K_1)}{n(N-1)}\sqrt{\eta_i\eta_j}\rho_{ij}(0),$$

$$\mathrm{Var}\left(I_1^{K_m}\right) \leq Cn^{-1/2}, \qquad \mathrm{Var}(I_3) \leq Cn^{-1/2}, \qquad \mathrm{Var}(R_1) \leq Cn^{-1},$$

$$\mathrm{Var}(R_3) \leq Cn^{-1}.$$

PROOF. Because $K_m > M$, $\varepsilon_i(t_r)$ and $\varepsilon_j(t_{r-K_m})$ are independent, so

$$E(I_3) = \sum_{m=1}^{N}\frac{a_m}{K_m}\sum_{K_m+1}^{n}E\left[\varepsilon_i(t_r)\varepsilon_j(t_{r-K_m})\right]$$

$$= \sum_{m=1}^{N}\frac{a_m}{K_m}\sum_{K_m+1}^{n}E\left[\varepsilon_i(t_r)\right]E\left[\varepsilon_j(t_{r-K_m})\right] = 0,$$

$$E(R_3) = \frac{K_N}{n(N-1)}\sum_{r=K_1+1}^{n}E\left[\varepsilon_i(t_r)\varepsilon_j(t_{r-K_1})\right]$$

$$= \frac{K_N}{n(N-1)}\sum_{r=K_1+1}^{n}E\left[\varepsilon_i(t_r)\right]E\left[\varepsilon_j(t_{r-K_1})\right] = 0.$$

For $I_1^{K_m}$ and R_1, we have

$$E(I_1^{K_m}) = \frac{1}{K_m} \sum_{r=1}^{K_m} E[\varepsilon_i(t_r)\varepsilon_j(t_r)] = \frac{1}{K_m} \sum_{r=1}^{K_m} \sqrt{\eta_i \eta_j} \rho_{ij}(0) = \sqrt{\eta_i \eta_j} \rho_{ij}(0),$$

$$ER_1 = \frac{K_N}{n(N-1)} \sum_{r=K_1+1}^{n} E[\varepsilon_i(t_r)\varepsilon_j(t_r)] = \frac{K_N(n-K_1)}{n(N-1)} \sqrt{\eta_i \eta_j} \rho_{ij}(0).$$

With the M-dependence of $\varepsilon_i(t_r)\varepsilon_j(t_r)$, we directly compute the variances of $I_1^{K_m}$ and R_1 as follows:

$$\mathrm{Var}(I_1^{K_m}) = \frac{1}{K_m^2} \sum_{r=1}^{K_m} \mathrm{Var}(\varepsilon_i(t_r)\varepsilon_j(t_r))$$

$$+ \frac{2}{K_m^2} \sum\sum_{1 \le r < r' \le K_m} \mathrm{Cov}(\varepsilon_i(t_r)\varepsilon_j(t_r), \varepsilon_i(t_{r'})\varepsilon_j(t_{r'}))$$

$$\le \frac{1}{K_m} \mathrm{Var}(\varepsilon_i(t_1)\varepsilon_j(t_1))$$

$$+ \frac{2}{K_m} \sum_{\ell=2}^{M+1} \mathrm{Cov}(\varepsilon_i(t_1)\varepsilon_j(t_1), \varepsilon_i(t_\ell)\varepsilon_j(t_\ell)) \le Cn^{-1/2},$$

$$\mathrm{Var}(R_1) = \left(\frac{K_N}{n(N-1)}\right)^2 \left[\sum_{r=K_1+1}^{n} \mathrm{Var}(\varepsilon_i(t_r)\varepsilon_j(t_r))\right.$$

$$\left. + 2 \sum\sum_{K_1+1 \le r < r' \le n} \mathrm{Cov}(\varepsilon_i(t_r)\varepsilon_j(t_r), \varepsilon_i(t_{r'})\varepsilon_j(t_{r'}))\right]$$

$$\le \left(\frac{K_N}{n(N-1)}\right)^2 \left[(n-K_1)\mathrm{Var}(\varepsilon_i(t_1)\varepsilon_j(t_1))\right.$$

$$\left. + 2(n-K_1) \sum_{\ell=2}^{M+1} \mathrm{Cov}(\varepsilon_i(t_1)\varepsilon_j(t_1), \varepsilon_i(t_\ell)\varepsilon_j(t_\ell))\right]$$

$$\le C/n.$$

We evaluate the variance of I_3 as follows:

$$E(I_3^2) = \sum_{m=1}^{N} \left(\frac{a_m}{K_m}\right)^2 E\left(\sum_{r=K_m+1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_{r-K_m})\right)^2$$

$$+ 2\sum\sum_{m<m'} \frac{a_m}{K_m} \frac{a_{m'}}{K_{m'}} E\left[\left(\sum_{r=K_m+1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_{r-K_m})\right)\right.$$

$$\times \left(\sum_{r'=K_{m'}+1}^{n} \varepsilon_i(t'_r)\varepsilon_j(t_{r'-K_{m'}}) \right) \Bigg]$$

$$= \sum_{m=1}^{N} \left(\frac{a_m}{K_m} \right)^2 \Bigg[\sum_{r=K_m+1}^{n} E\big(\varepsilon_i^2(t_r)\varepsilon_j^2(t_{r-K_m})\big)$$

$$+ 2 \sum_{r<r'} \sum E\big(\varepsilon_i(t_r)\varepsilon_j(t_{r-K_m})\varepsilon_i(t_{r'})\varepsilon_j(t_{r'-K_m})\big) \Bigg]$$

$$+ 2 \sum_{m<m'} \sum \frac{a_m}{K_m} \frac{a_{m'}}{K_{m'}} \Bigg[\sum_{r=K_{m'}+1}^{n} E\big(\varepsilon_i^2(t_r)\varepsilon_j(t_{r-K_m})\varepsilon_j(t_{r-K_{m'}})\big)$$

$$+ \sum_{r=1}^{n-K_{m'}} E\big(\varepsilon_j^2(t_r)\varepsilon_i(t_{r+K_m})\varepsilon_i(t_{r+K_{m'}})\big)$$

$$+ 2 \sum_{r<r'} \sum E\big(\varepsilon_i(t_r)\varepsilon_j(t_{r-K_m})\varepsilon_i(t_{r'})\varepsilon_j(t_{r'-K_{m'}})\big) \Bigg]$$

$$= \sum_{m=1}^{N} \left(\frac{a_m}{K_m} \right)^2 \Bigg[(n-K_m)\eta_i\eta_j$$

$$+ 2 \sum_{\ell=2}^{(n-K_m)\wedge(M+1)} (n-K_m-\ell+1)E\big(\varepsilon_i(t_1)\varepsilon_i(t_\ell)\big)$$

$$\times E\big(\varepsilon_j(t_1)\varepsilon_j(t_\ell)\big) \Bigg]$$

$$+ 2 \sum_{m<m'<m+M+1} \sum \frac{a_m}{K_m} \frac{a_{m'}}{K_{m'}} \Bigg[(n-K_{m'})\eta_i E\big(\varepsilon_j(t_1)\varepsilon_j(t_{K_{m'}-K_m+1})\big)$$

$$+ (n-K_m)\eta_j E\big(\varepsilon_i(t_n)\varepsilon_i(t_{n-K_{m'}+K_m})\big)$$

$$+ 2 \sum_{\ell=2}^{(n-K_{m'})\wedge(M+1)} (n-K_{m'}-\ell+1)E\big(\varepsilon_i(t_1)\varepsilon_i(t_\ell)\big)$$

$$\times E\big(\varepsilon_j(t_1)\varepsilon_j(t_{\ell+K_{m'}-K_m})\big) \Bigg]$$

$$\leq C_1\eta_i\eta_j N(1/N^2)^2(n-K_1) + C_2\eta_i\eta_j(1/N^2)^2(n-K_1) \asymp Cn^{-1/2},$$

where the inequality is from the fact that the M-dependence of $(\varepsilon_1(t_\ell), \ldots, \varepsilon_p(t_\ell))$ implies zero expectations of $\varepsilon_i(\cdot)\varepsilon_j(\cdot)$ for lags larger than M.

Similarly, we have

$$
\begin{aligned}
E(R_3^2) &= \left(\frac{K_N}{n(N-1)}\right)^2 E\left(\sum_{r=K_1+1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_{r-K_1})\right)^2 \\
&= \left(\frac{K_N}{n(N-1)}\right)^2\left[\sum_{r=K_1+1}^{n} E\big(\varepsilon_i^2(t_r)\varepsilon_j^2(t_{r-K_1})\big)\right.
\end{aligned}
$$

$$
\left. + 2\sum_{r<r'}\sum E\big(\varepsilon_i(t_r)\varepsilon_j(t_{r-K_1})\varepsilon_i(t_{r'})\varepsilon_j(t_{r'-K_1})\big)\right]
$$

$$
= \left(\frac{K_N}{n(N-1)}\right)^2\left[(n-K_1)\eta_i\eta_j\right.
$$

$$
+ 2\sum_{\ell=2}^{(n-K_1)\wedge(M+1)} (n-K_1-\ell+1)E\big(\varepsilon_i(t_1)\varepsilon_i(t_\ell)\big)
$$

$$
\left. \times E\big(\varepsilon_j(t_1)\varepsilon_j(t_\ell)\big)\right]
$$

$$
\leq C_1\eta_i\eta_j(1/n)^2(n-K_1) \asymp C_2/n,
$$

where the inequality is from the fact that the M-dependence of $(\varepsilon_1(t_\ell), \ldots, \varepsilon_p(t_\ell))$ implies zero expectations of $\varepsilon_i(\cdot)\varepsilon_j(\cdot)$ for lags larger than M. $\quad\square$

APPENDIX II: PROOF OF PROPOSITION 9

We break the proof into a few major technical lemmas which are proved in Sections II.2–II.3. Without loss of generality we consider only the case $i = 1$ and prove that there exists a constant $C_1 > 0$ such that $\|\bar{\mathbb{P}}_{1,0} \wedge \bar{\mathbb{P}}_{1,1}\| \geq C_1$.

The following lemma turns the problem of bounding the total variation affinity into a chi-square distance calculation. Denote the projection of $\theta \in \Theta$ to $\boldsymbol{\Gamma}$ by $\gamma(\theta) = (\gamma_i(\theta))_{1\leq i\leq r}$ and to Λ by $\lambda(\theta) = (\lambda_i(\theta))_{1\leq i\leq r}$. More generally, for a subset $A \subseteq \{1, 2, \ldots, r\}$, we define a projection of θ to a subset of $\boldsymbol{\Gamma}$ by $\gamma_A(\theta) = (\gamma_i(\theta))_{i\in A}$. A particularly useful example of set A is $\{1, \ldots, i-1, i+1, \ldots, r\}$ for which we use $\gamma_{-i}(\theta) = (\gamma_1(\theta), \ldots, \gamma_{i-1}(\theta), \gamma_{i+1}(\theta), \gamma_r(\theta))$. $\lambda_A(\theta)$ and $\lambda_{-i}(\theta)$ are defined similarly. We define the set $\Lambda_A = \{\lambda_A(\theta) : \theta \in \Theta\}$. For $a \in \{0, 1\}$, $b \in \{0, 1\}^{r-1}$, and $c \in \Lambda_{-i} \subseteq B^{r-1}$, let

$$
\Theta_{(i,a,b,c)} = \{\theta \in \Theta : \gamma_i(\theta) = a, \gamma_{-i}(\theta) = b \text{ and } \lambda_{-i}(\theta) = c\},
$$

and $D_{(i,a,b,c)} = \text{Card}(\Theta_{(i,a,b,c)})$ which depends actually on the value of c, not values of i, a and b for the parameter space Θ constructed in Section 6.2. Define

the mixture distribution

$$
\text{(70)} \qquad \bar{\mathbb{P}}_{(i,a,b,c)} = \frac{1}{D_{(i,a,b,c)}} \sum_{\theta \in \Theta_{(i,a,b,c)}} \mathbb{P}_\theta.
$$

In other words, $\bar{\mathbb{P}}_{(i,a,b,c)}$ is the mixture distribution over all $\mathbb{P}_\theta$ with $\lambda_i(\theta)$ varying over all possible values while all other components of θ remain fixed. Define

$$
\Theta_{-1} = \{(b,c) : \text{there exists a } \theta \in \Theta \text{ such that } \gamma_{-1}(\theta) = b \text{ and } \lambda_{-1}(\theta) = c\}.
$$

LEMMA 15.   *If there is a constant $C_2 < 1$ such that*

$$
\text{(71)} \qquad \underset{(\gamma_{-1}, \lambda_{-1}) \in \Theta_{-1}}{\text{Average}} \left\{ \int \left(\frac{d\bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})}}{d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}} \right)^2 d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})} - 1 \right\} \le C_2^2,
$$

then $\|\bar{\mathbb{P}}_{1,0} \wedge \bar{\mathbb{P}}_{1,1}\| \ge 1 - C_2 > 0$.

We can prove Lemma 15 using the same arguments as the proof of Lemma 8 in Cai and Zhou (2012). To complete the proof of Proposition 9 we need to verify only equation (71).

II.1. Technical lemmas for proving equation (71).   From the definition of $\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}$ in equation (70) and $\theta = (\gamma, \lambda)$ with $\gamma = (\gamma_1, \ldots, \gamma_r)$ and $\lambda = (\lambda_1, \ldots, \lambda_r)$, $\gamma_1 = 0$ implies $\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}$ is a product of n multivariate normal distributions each with a covariance matrix,

$$
\text{(72)} \qquad \Sigma_{l,0} = \begin{pmatrix} 1 & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & \mathbf{S}_{(p-1) \times (p-1)} \end{pmatrix} + (a_l - 1)\mathbf{I}_p \qquad \text{for } l = 1, 2, \ldots, n,
$$

where $\mathbf{S}_{(p-1) \times (p-1)} = (s_{ij})_{2 \le i, j \le p}$ is uniquely determined by $(\gamma_{-1}, \lambda_{-1}) = ((\gamma_2, \ldots, \gamma_r), (\lambda_2, \ldots, \lambda_r))$ with

$$
s_{ij} = \begin{cases} 1, & i = j, \\ \epsilon_{n,p}, & \gamma_i = \lambda_i(j) = 1, \\ 0, & \text{otherwise.} \end{cases}
$$

Let $n_{\lambda_{-1}}$ be the number of columns of λ_{-1} with column sum equal to $2k$ and $p_{\lambda_{-1}} = r - n_{\lambda_{-1}}$. Since $n_{\lambda_{-1}} \cdot 2k \le r \cdot k$, the total number of 1s in the upper triangular matrix, we have $n_{\lambda_{-1}} \le r/2$, which implies $p_{\lambda_{-1}} = r - n_{\lambda_{-1}} \ge r/2 \ge p/4 - 1$. From equations (70) and $\theta = (\gamma, \lambda)$ with $\gamma = (\gamma_1, \ldots, \gamma_r)$ and $\lambda = (\lambda_1, \ldots, \lambda_r)$, $\bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})}$ is an average of $\binom{p_{\lambda_{-1}}}{k}$ number of products of multivariate normal distributions each with covariance matrix of the following form:

$$
\text{(73)} \qquad \begin{pmatrix} 1 & \mathbf{r}_{1 \times (p-1)} \\ \mathbf{r}_{(p-1) \times 1} & \mathbf{S}_{(p-1) \times (p-1)} \end{pmatrix} + (a_l - 1)\mathbf{I}_p \qquad \text{for } l = 1, 2, \ldots, n,
$$

where $\|\mathbf{r}\|_0 = k$ with nonzero elements of r equal to $\epsilon_{n,p}$ and the submatrix $\mathbf{S}_{(p-1) \times (p-1)}$ is the same as the one for $\Sigma_{l,0}$ given in (72). Note that the indices γ_i and λ_i are dropped from $\mathbf{r}$ and $\mathbf{S}$ to simplify the notation.

With Lemma 15 in place, it remains to establish equation (71) in order to prove Proposition 9. The following lemma is useful for calculating the cross product terms in the chi-square distance between Gaussian mixtures. The proof of the lemma is straightforward and is thus omitted.

LEMMA 16.   *Let g_i be the density function of $N(0, \Sigma_i)$ for $i = 0, 1$ and 2, respectively. Then*

$$\int \frac{g_1 g_2}{g_0} = \frac{1}{[\det(\mathbf{I} - \Sigma_0^{-2}(\Sigma_1 - \Sigma_0)(\Sigma_2 - \Sigma_0))]^{1/2}}.$$

Let $\Sigma_{l,i}$, $i = 1$ or 2, be two covariance matrices of the form (73). Note that $\Sigma_{l,i}$, $i = 0, 1$ or 2, differs from each other only in the first row/column. Then $\Sigma_{l,i} - \Sigma_{l,0}$, $i = 1$ or 2, has a very simple structure. The nonzero elements only appear in the first row/column, and in total there are $2k$ nonzero elements. This property immediately implies the following lemma which makes the problem of studying the determinant in Lemma 16 relatively easy.

LEMMA 17.   *Let $\Sigma_{l,i}$, $i = 1$ and 2, be matrices of the form (73). Define J to be the number of overlapping $\epsilon_{n,p}$'s between $\Sigma_{l,1}$ and $\Sigma_{l,2}$ on the first row, and*

$$Q \overset{\triangle}{=} (q_{ij})_{1 \leq i, j \leq p} = (\Sigma_{l,1} - \Sigma_{l,0})(\Sigma_{l,2} - \Sigma_{l,0}).$$

There are index subsets I_r and I_c in $\{1, 2, \ldots, p\}$ with $\mathrm{Card}(I_r) = \mathrm{Card}(I_c) = k$ and $\mathrm{Card}(I_r \cap I_c) = J$ such that

$$q_{ij} = \begin{cases} J\epsilon_{n,p}^2, & i = j = 1, \\ \epsilon_{n,p}^2, & i \in I_r \text{ and } j \in I_c, \\ 0, & \text{otherwise}, \end{cases}$$

and the matrix $(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})$ has rank 2 with two identical nonzero eigenvalues $J\epsilon_{n,p}^2$ when $J > 0$.

Let

$$(74) \qquad R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} = -\log \det(I - \Sigma_{l,0}^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})),$$

where $\Sigma_{l,0}$ is defined in (72) and determined by $(\gamma_{-1}, \lambda_{-1})$, and $\Sigma_{l,1}$ and $\Sigma_{l,2}$ have the first row λ_1 and λ_1', respectively. We drop the indices λ_1, λ_1' and $(\gamma_{-1}, \lambda_{-1})$ from Σ_i to simplify the notation. Define

$$\Theta_{-1}(a_1, a_2) = \big\{(b, c) : \text{there exist } \theta_i \in \Theta, i = 1, 2, , \text{ such that } \lambda_1(\theta_i) = a_i$$

$$\text{and } \lambda_{-1}(\theta_i) = c\big\}.$$

It is a subset of Θ_{-1} in which the element can pick both a_1 and a_2 as the first row to form parameters in Θ. From Lemma 16 the left-hand side of equation (71) can be written as

$$
\begin{aligned}
(75) \quad & \underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}}{\text{Average}} \left\{ \underset{\lambda_1,\lambda_1'\in\Lambda_1(\lambda_{-1})}{\text{Average}} \left[\exp\left(\frac{1}{2}\sum_{l=1}^{n} R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}\right) - 1 \right] \right\} \\
& = \text{Average}\left\{ \underset{\substack{\lambda_1,\lambda_1'\in B}}{} \underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\lambda_1')}{\text{Average}} \left[\exp\left(\frac{1}{2}\sum_{l=1}^{n} R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}\right) - 1 \right] \right\},
\end{aligned}
$$

where B is defined in step 1.

Lemmas 17 and 18 below show that $R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}$ is approximately equal to

$$
- \log\det(I - a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})) = -2\log(1 - a_l^{-2} J\epsilon_{n,p}^2).
$$

Define

$$
\Lambda_{1,J} = \big\{(\lambda_1, \lambda_1') \in \Lambda_1 \otimes \Lambda_1 : \text{the number of overlapping } \epsilon_{n,p}\text{'s between } \lambda_1
$$
$$
\text{and } \lambda_1' \text{ is } J\big\}.
$$

LEMMA 18. *For $R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}$ defined in equation (74), we have*

$$
(76) \qquad R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} = -2\log(1 - Ja_l^{-2}\epsilon_{n,p}^2) + \delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}},
$$

where $\delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}$ satisfies

$$
(77) \qquad \underset{(\lambda_1,\lambda_1')\in\Lambda_{1,J}}{\text{Average}} \left[\underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\lambda_1')}{\text{Average}} \exp\left(\frac{1}{2}\sum_{l=1}^{n} \delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}\right) \right] \le 3/2,
$$

uniformly over all J defined in Lemma 17.

We will prove Lemma 18 in Section II.3.

II.2. Proof of equation (71). We are now ready to establish equation (71) using Lemma 18. It follows from equation (76) in Lemma 18 that

$$
\underset{\lambda_1,\lambda_1'\in B}{\text{Average}}\left\{ \underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\lambda_1')}{\text{Average}} \left[\exp\left(\frac{1}{2}\sum_{l=1}^{n} R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}\right) - 1 \right] \right\}
$$
$$
= \text{Average}_J\left\{ -\sum_{l=1}^{n} \log\left(1 - \frac{J\epsilon_{n,p}^2}{a_l^2}\right) \right.
$$
$$
\left. \times \underset{(\lambda_1,\lambda_1')\in\Lambda_{1,J}}{\text{Average}} \left[\underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\lambda_1')}{\text{Average}} \exp\left(\frac{1}{2}\sum_{l=1}^{n} \delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}\right) \right] - 1 \right\}.
$$

Recall that J is the number of overlapping $\epsilon_{n,p}$'s between $\Sigma_{l,1}$ and $\Sigma_{l,2}$ on the first row. It is easy to see that J has the hypergeometric distribution with

$$\mathbb{P}(\text{number of overlapping } \epsilon_{n,p}\text{'s} = J)$$

(78)
$$= \binom{k}{J} \binom{p_{\lambda_{-1}} - k}{k - J} \bigg/ \binom{p_{\lambda_{-1}}}{k}$$

$$\leq \left(\frac{k^2}{p_{\lambda_{-1}} - k} \right)^J.$$

Equations (77) and (78) imply

$$\underset{(\gamma_{-1}, \lambda_{-1}) \in \Theta_{-1}}{\text{Average}} \left\{ \int \left(\frac{d\bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})}}{d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}} \right)^2 d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})} - 1 \right\}$$

$$\leq \sum_{J \geq 0} \frac{\binom{k}{J} \binom{p_{\lambda_{-1}} - k}{k - J}}{\binom{p_{\lambda_{-1}}}{k}} \left\{ -\sum_{l=1}^{n} \log(1 - J\epsilon_{n,p}^2 / a_l^2) \right\}^{\frac{3}{2}} - 1$$

$$\leq C \sum_{J \geq 1} (p^{(\beta - 1)/\beta})^{-J} \exp\left(2J \sum_{l=1}^{n} a_l^{-2} \cdot \frac{\upsilon^2 \log p}{\sqrt{n}} \right) + 1/2$$

$$\leq C \sum_{J \geq 1} (p^{(\beta - 1)/\beta})^{-J} \exp\left(2J c_\kappa \sqrt{n} \cdot \frac{\upsilon^2 \log p}{\sqrt{n}} \right) + 1/2$$

$$\leq C \sum_{J \geq 1} (p^{(\beta - 1)/\beta})^{-J} \exp(2c_\kappa J \upsilon^2 \log p) + 1/2$$

$$\leq C \sum_{J \geq 1} (p^{(\beta - 1)/(2\beta)})^{-J} + 1/2 < C_2^2,$$

where the third inequality is from (37), the fifth inequality is due to (36) and the last inequality is obtained by setting $C_2^2 = 3/4$.

II.3. Proof of Lemma 18. Define

(79)
$$A_l = \left[I - a_l^{-2} (\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2}) \right]^{-1}$$
$$\times \left(a_l^2 (\Sigma_{l,0})^{-2} - I \right) a_l^{-2} (\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})$$

and

$$\delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1}, \lambda_{-1}} = -\log \det(I - A_l).$$

We rewrite $R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}$ as follows:

$$
\begin{aligned}
R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} &= -\log\det[I - a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2}) \\
&\qquad\qquad - (a_l^2\Sigma_{l,0}^{-2} - I)a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})]
\end{aligned}
$$

$$
\begin{aligned}
(80) \qquad &= -\log\det\{[I - A_l] \cdot [I - a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})]\} \\
&= -\log\det[I - a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})] - \log\det(I - A_l) \\
&= -2\log(1 - J\epsilon_{n,p}^2/a_l^2) + \delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}},
\end{aligned}
$$

where the last equation follows from Lemma 17.

Now we are ready to establish equation (77). For simplicity we will write matrix norm $\|\cdot\|_2$ as $\|\cdot\|$ below. It is important to observe that $\mathrm{rank}(A_l) \leq 2$ due to the simple structure of $(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})$. Let ϱ_l be an eigenvalue of A_l. It is easy to see that

$$
\begin{aligned}
(81) \qquad |\varrho_l| &\leq \|A_l\| \\
&\leq \|a_l^2\Sigma_{l,0}^{-2} - I\| \cdot a_l^{-2}\|\Sigma_{l,0} - \Sigma_{l,1}\|\|\Sigma_{l,0} - \Sigma_{l,2}\| \\
&\qquad /(1 - a_l^{-2}\|\Sigma_{l,0} - \Sigma_{l,1}\|\|\Sigma_{l,0} - \Sigma_{l,2}\|) \\
&\leq \left(\left(\frac{3}{2}\right)^2 - 1\right)\frac{1}{3} \cdot \frac{1}{3}\bigg/\left(1 - \frac{1}{3} \cdot \frac{1}{3}\right) = 5/32 < 1/6,
\end{aligned}
$$

since $\|a_l^{-1}(\Sigma_{l,0} - \Sigma_{l,1})\| \leq \|a_l^{-1}(\Sigma_{l,0} - \Sigma_{l,1})\|_1 = 2k\epsilon_{n,p} < 1/3$ and $\lambda_{\min}(a_l^{-1}\Sigma_{l,0}) \geq 1 - \|I - a_l^{-1}\Sigma_{l,0}\| \geq 1 - \|I - a_l^{-1}\Sigma_{l,0}\|_1 > 2/3$ from equation (38).

Note that (81) and

$$
|\log(1 - x)| \leq 2|x|, \qquad \text{for } |x| < 1/6,
$$

imply

$$
\delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} \leq 4\|A_l\|,
$$

and then

$$
(82) \qquad \exp\left(\frac{1}{2}\sum_{l=1}^{n}\delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}\right) \leq \exp\left(2\sum_{l=1}^{n}\|A_l\|\right).
$$

Since

$$
(83) \qquad \begin{cases} \|I - a_l^{-1}\Sigma_{l,0}\| \leq \|I - a_l^{-1}\Sigma_{l,0}\|_1 = 2k\epsilon_{n,p} < 1/3 < 1, \\ \|a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})\| \leq \frac{1}{3} \cdot \frac{1}{3} < 1, \end{cases}
$$

we write

$$a_l^2 \Sigma_{l,0}^{-2} - I = \left(I - \left(I - a_l^{-1}\Sigma_{l,0}\right)\right)^{-2} - I$$

(84)
$$= \left(I + \sum_{k=1}^{\infty} \left(I - a_l^{-1}\Sigma_{l,0}\right)^k\right)^2 - I$$

$$= \left[\sum_{m=0}^{\infty} (m+2)\left(I - a_l^{-1}\Sigma_{l,0}\right)^m\right]\left(I - a_l^{-1}\Sigma_{l,0}\right),$$

where

(85)
$$\left\|\sum_{m=0}^{\infty} (m+2)\left(I - a_l^{-1}\Sigma_{l,0}\right)^m\right\| \le \sum_{m=0}^{\infty} (m+2)\left(\frac{1}{3}\right)^m < 3.$$

Define

(86)
$$A_{l*} = \left(I - a_l^{-1}\Sigma_{l,0}\right) \cdot a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2}).$$

From equations (79) and (83)–(86) we have

$$\|A_l\| \le \left\|\left[I - a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})\right]^{-1}\right\|$$

$$\times \left\|\sum_{m=0}^{\infty} (m+2)\left(I - a_l^{-1}\Sigma_{l,0}\right)^m\right\| \|A_{l*}\|$$

$$< \frac{1}{1 - (1/3) \cdot (1/3)} \cdot 3 \cdot \|A_{l*}\| = \frac{27}{8}\|A_{l*}\| \le \frac{27}{8}\max\{\|A_{l*}\|_1, \|A_{l*}\|_\infty\}.$$

The above result and (82) indicate that the proof of Lemma 18 is complete if we show

(87)
$$\operatorname*{Average}_{(\lambda_1,\lambda_1') \in \Lambda_{1,J}} \left[\operatorname*{Average}_{(\gamma_{-1},\lambda_{-1}) \in \Theta_{-1}(\lambda_1,\lambda_1')}\right.$$

$$\left. \times \exp\left(\frac{27}{2}\sum_{l=1}^{n}\max\{\|A_{l*}\|_1, \|A_{l*}\|_\infty\}\right)\right] \le 3/2,$$

where $\|A_{l*}\|_1$ and $\|A_{l*}\|_\infty$ depend on the values of λ_1, λ_1' and $(\gamma_{-1}, \lambda_{-1})$. We dropped the indices λ_1, λ_1' and $(\gamma_{-1}, \lambda_{-1})$ from A_l to simplify the notation.

Let $E_m = \{1, 2, \ldots, r\}/\{1, m\}$. Let $n_{\lambda_{E_m}}$ be the number of columns of λ_{E_m} with column sum at least $2k - 2$ for which two rows cannot freely take value 0 or 1 in this column. Then we have $p_{\lambda_{E_m}} = r - n_{\lambda_{E_m}}$. Without loss of generality we assume that $k \ge 3$. Since $n_{\lambda_{E_m}} \cdot (2k - 2) \le r \cdot k$, the total number of 1s in the upper triangular matrix by the construction of the parameter set, we thus have

$n_{\lambda_{E_m}} \leq r \cdot \frac{3}{4}$, which immediately implies $p_{\lambda_{E_m}} = r - n_{\lambda_{E_m}} \geq \frac{r}{4} \geq p/8 - 1$. Thus we have for every nonnegative integer t,

$$\mathbb{P}\big(\max\{\|A_{l*}\|_1, \|A_{l*}\|_\infty\} \geq 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 \cdot a_l^{-3}\big)$$

$$\leq \mathbb{P}\big(\|A_{l*}\|_1 \geq 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 \cdot a_l^{-3}\big) + \mathbb{P}\big(\|A_{l*}\|_\infty \geq 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 \cdot a_l^{-3}\big)$$

$$\leq 2 \sum_m \underset{\lambda_{E_m}}{\text{Average}} \frac{\binom{k}{t}\binom{p_{\lambda_{E_m}}}{k-t}}{\binom{p_{\lambda_{E_m}}}{k}} \leq 2p\left(\frac{k^2}{p/8 - 1 - k}\right)^t$$

from equation (78), which immediately implies

$$\underset{(\lambda_1, \lambda_1') \in \Lambda_{1,J}}{\text{Average}} \left[\underset{(\gamma_{-1}, \lambda_{-1}) \in \Theta_{-1}(\lambda_1, \lambda_1')}{\text{Average}} \exp\left(\frac{27}{2} \sum_{l=1}^n \max\{\|A_{l*}\|_1, \|A_{l*}\|_\infty\}\right) \right]$$

$$\leq \exp\left(\frac{27}{2} \sum_{l=1}^n \frac{4\beta}{\beta - 1} \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 \cdot a_l^{-3}\right)$$

$$+ \int_{2\beta/(\beta-1)}^\infty \left(27k\epsilon_{n,p}^3 \sum_{l=1}^n a_l^{-3}\right)$$

$$\text{(88)} \qquad\qquad \times \exp\left(\frac{27}{2} \sum_{l=1}^n 2t \cdot \epsilon_{n,p} \cdot k\epsilon_{n,p}^2 \cdot a_l^{-3}\right) 2p\left(\frac{k^2}{p/8 - 1 - k}\right)^{t-1} dt$$

$$\leq \exp\left(54 \cdot \left(\sum_{l=1}^n a_l^{-3}\right) \cdot \frac{\beta}{\beta - 1} \cdot k\epsilon_{n,p}^3\right)$$

$$+ 2p \int_{2\beta/(\beta-1)}^\infty \exp\left[(t+1) \cdot 27 \left(\sum_{l=1}^n a_l^{-3}\right) k\epsilon_{n,p}^3 \right.$$

$$\left. - (t-1) \log \frac{p/8 - 1 - k}{k^2} \right] dt.$$

Note that (37) implies

$$\sum_{l=1}^n a_l^{-3} \leq \sum_{l=1}^n a_l^{-2} \leq c_K \sqrt{n},$$

using (14) and (34) we have

$$2\sqrt{n}k\epsilon_{n,p}^3 \leq \sqrt{n}\pi_n(p)\epsilon_{n,p}^{3-q}$$

$$\leq \aleph v^{3-q} n^{1/2} n^{(1-q)/4} (\log p)^{(q-3)/2} n^{(q-3)/4} (\log p)^{(3-q)/2}$$

$$= \aleph v^{3-q},$$

and thus we can bound the first term on the right-hand side of (88),

$$\exp\left(54 \cdot c_\kappa \sqrt{n} \cdot \frac{\beta}{\beta - 1} \cdot k\epsilon_{n,p}^3\right) \le \exp\left(\frac{\beta}{\beta - 1} \cdot 27 c_\kappa v^2 \cdot \aleph v^{1-q}\right) \le \exp(1/3) < 3/2,$$

where the second inequality is from (35) and (36). We will show that the second term on the right-hand side of (88) is negligible and hence establish (87). Indeed, since we have just shown that

$$27\left(\sum_{l=1}^n a_l^{-3}\right)k\epsilon_{n,p}^3 \le \frac{\beta - 1}{6\beta},$$

the second term on the right-hand side of (88) is bounded by

$$2p \int_{2\beta/(\beta-1)}^\infty \exp\left[(t+1)\frac{\beta-1}{6\beta} - (t-1)\log\frac{p/8 - 1 - k}{k^2}\right]dt$$

$$= 2\left(\log\frac{p/8 - 1 - k}{k^2} - \frac{\beta-1}{6\beta}\right)^{-1}$$

$$\times \exp\left[\log p + \left(\frac{2\beta}{\beta-1} + 1\right)\frac{\beta-1}{6\beta} - \left(\frac{2\beta}{\beta-1} - 1\right)\log\frac{p/8 - 1 - k}{k^2}\right]$$

$$= O\left(p^{-1/\beta}[\log p]^{6/(\beta-1)+2}\right) = o(1),$$

where the second equality is from the fact that (14) and (34) together with $p \ge n^{\beta/2}$ indicate

$$k^2 \le \pi_n(p)\epsilon_{n,p}^{-2q}/4 \le \frac{\aleph v^{-2q}\sqrt{n}}{4\log^3 p} \le \frac{\aleph v^{-2q} p^{1/\beta}}{4\log^3 p},$$

and then

$$\left(\frac{2\beta}{\beta-1} - 1\right)\log\frac{p/8 - k}{k^2}$$

$$= \left(\frac{2\beta}{\beta-1} - 1\right)\log(pk^{-2})[1 + o(1)]$$

$$\ge \left(\frac{2\beta}{\beta-1} - 1\right)\left[\frac{\beta-1}{\beta}\log p + 3\log\log p - \log(Mv^{-2q}/4)\right]$$

$$= \left(1 + \frac{1}{\beta}\right)\log p[1 + o(1)].$$

REFERENCES

AïT-SAHALIA, Y., FAN, J. and XIU, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.* **105** 1504–1517. MR2796567

AÏT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18** 351–416.

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536. MR2468558

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162** 149–169. MR2795610

BIBINGER, M. and REIß, M (2011). Spectral estimation of covolatility from noisy observations using local weights. Preprint, Humboldt-Universität zu Berlin.

BICKEL, P. J. and LEVINA, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

BICKEL, P. J. and LEVINA, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885

CAI, T. and ZHOU, H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420.

CALLEN, J., GOVINDARAJ, S. and XU, L. (2000). Large time and small noise asymptotic results for mean reverting diffusion processes with applications. *Econom. Theory* **16** 401–419. MR1778742

CHRISTENSEN, K., KINNEBROCK, S. and PODOLSKIJ, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics* **159** 116–133. MR2720847

CURCI, G. and CORSI, F. (2012). Discrete sine transform for multi-scale realized volatility measures. *Quant. Finance* **12** 263–279. MR2881619

EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. MR2485011

FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991

FAN, J., LI, Y. and YU, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Amer. Statist. Assoc.* **107** 412–428. MR2949370

FAN, J. and WANG, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *J. Amer. Statist. Assoc.* **102** 1349–1362. MR2372538

GLOTER, A. and JACOD, J. (2001a). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Stat.* **5** 225–242 (electronic). MR1875672

GLOTER, A. and JACOD, J. (2001b). Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Stat.* **5** 243-260.

HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. MR2277742

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. and VETTER, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119** 2249–2276. MR2531091

JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448

KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940

LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459

LEDOIT, O. and WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88** 365–411. MR2026339

MUESCHKE, N. J. and ANDREWS, M. J. (2006). Investigation of scalar measurement error in dif-
    fusion and mixing processes. *Experiments in Fluids* **40** 165–175.
MUNK, A. and SCHMIDT-HIEBER, J. (2010). Lower bounds for volatility estimation in microstruc-
    ture noise models. In *Borrowing Strength*: *Theory Powering Applications—A Festschrift for
    Lawrence D. Brown*. *Inst. Math. Stat. Collect.* **6** 43–55. IMS, Beachwood, OH. MR2798510
REIß, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations.
    Ann. Statist. **39** 772–802. MR2816338
SALKUYEH, D. K. (2006). Positive integer powers of the tridiagonal Toeplitz matrices. *Int. Math.
    Forum* **1** 1061–1065. MR2256665
SAULIS, L. and STATULEVIČIUS, V. A. (1991). *Limit Theorems for Large Deviations. Mathematics
    and Its Applications* (*Soviet Series*) **73**. Kluwer Academic, Dordrecht. MR1171883
TAO, M., WANG, Y., YAO, Q. and ZOU, J. (2011). Large volatility matrix inference via com-
    bining low-frequency and high-frequency approaches. *J. Amer. Statist. Assoc.* **106** 1025–1040.
    MR2894761
THORIN, G. O. (1948). Convexity theorems generalizing those of M. Riesz and Hadamard with
    some applications. *Medd. Lunds Univ. Mat. Sem.* **9** 1–58. MR0025529
WANG, Y. and ZOU, J. (2010). Vast volatility matrix estimation for high-frequency financial data.
    Ann. Statist. **38** 943–978. MR2604708
WHITMORE, G. A. (1995). Estimating degradation by a Wiener diffusion process subject to mea-
    surement error. *Lifetime Data Anal.* **1** 307–319.
WILKINSON, J. H. (1988). *The Algebraic Eigenvalue Problem*. Oxford Univ. Press, New York.
    MR0950175
WU, W. B. and POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices
    of longitudinal data. *Biometrika* **90** 831–844. MR2024760
XIU, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data.
    J. Econometrics **159** 235–250. MR2720855
YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model.
    Biometrika **94** 19–35. MR2367824
ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-
    scale approach. *Bernoulli* **12** 1019–1043. MR2274854
ZHANG, L. (2011). Estimating covariation: Epps effect, microstructure noise. *J. Econometrics* **160**
    33–47. MR2745865
ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determin-
    ing integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411.
    MR2236450
ZHENG, X. and LI, Y. (2011). On the estimation of integrated covariance matrices of high dimen-
    sional diffusion processes. *Ann. Statist.* **39** 3121–3151. MR3012403

M. TAO   H. H. ZHOU
Y. WANG   DEPARTMENT OF STATISTICS
DEPARTMENT OF STATISTICS   YALE UNIVERSITY
UNIVERSITY OF WISCONSIN-MADISON   NEW HAVEN, CONNECTICUT 06511
MADISON, WISCONSIN 53706   USA
USA   E-MAIL: huibin.zhou@yale.edu
E-MAIL: minjing@stat.wisc.edu
        yzwang@stat.wisc.edu