Optimal Volatility Matrix Estimation for High Dimensional Diffusions With Noise Contamination

Minjing Tao and Yazhen Wang
University of Wisconsin-Madison
Harrison H. Zhou
Yale University

October 31, 2011

Abstract

Large matrix estimation gains an increasing attention in recent years. This paper investigates the high dimensional statistical problem where a p-dimensional diffusion process is observed with measurement errors at n distinct time points, and our goal is to estimate the volatility matrix of the diffusion process. We establish the minimax theory for estimating large sparse volatility matrices under matrix spectral norm as both n and p go to infinity. The theory shows that the optimal convergence rate depends on n and p through $n^{-1/4}\sqrt{\log p}$ and a volatility matrix estimator is explicitly constructed to achieve the optimal convergence rate.

Key words and phrases: Diffusion process, large matrix estimation, measurement error, minimax lower bound, optimal convergence rate, small n and large p, thresholding, volatility matrix estimator.

AMS 2000 subject classifications: Primary 62G05, 62H12; secondary 62M05

1 Introduction

Diffusions are widely employed in modern scientific studies in fields ranging from biology and finance to engineering and physical science. The diffusion models play a key role in describing complex dynamic systems where it is essential to incorporate internally or externally originating random fluctuations in the system. See for example, Fan and Wang (2007), Mueschke and Andrews (2006), Prakasa Rao (1999), Wang and Zou (2010), Whitmore (1995) and Zhang et. al. (2005). The scientific studies motivate this paper to investigate optimal estimation of large matrices for high dimensional diffusions with noise contamination.

Consider process $\mathbf{X}(t) = (X_1(t), \cdots, X_p(t))^T$ following the continuous-time diffusion model

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

where $\boldsymbol{\mu}_t$ is a p-dimensional drift vector, $\mathbf{B}_t$ is a p-dimensional standard Brownian motion, and $\boldsymbol{\sigma}_t$ is a p by p matrix. Continuous-time process $\mathbf{X}_t$ is observed only at discrete time points with measurement errors, that is, the observed discrete data $Y_i(t_\ell)$ obey

$$Y_i(t_\ell) = X_i(t_\ell) + \varepsilon_i(t_\ell), \quad i = 1, \cdots, p, \ t_\ell = \ell/n, \ \ell = 1, \cdots, n, \tag{2}$$

where $\varepsilon_i(t_\ell)$ are independent 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 $Y_i(t_\ell)$, $i = 1, \cdots, p$, $\ell = 1, \cdots, n$.

Below is the main theorem of the paper that establishes the minimax risk for estimating $\boldsymbol{\Gamma}$ based on $Y_i(t_\ell)$, $i = 1, \cdots, p$, $\ell = 1, \cdots, n$, which is a consequence of Theorems 3 and 4 in Sections 2 and 3.

Theorem 1 *For models (1)-(2) under the conditions A1-A3 specified in Section 2, we have that as $n, p \to \infty$,*

$$C_* \left[\pi_n(p) \left(n^{-1/4} \sqrt{\log p} \right)^{1-q} \right]^2 \le \inf_{\check{\boldsymbol{\Gamma}}} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E} \left\| \check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma} \right\|_d^2 \le C^* \left[\pi_n(p) \left(n^{-1/4} \sqrt{\log p} \right)^{1-q} \right]^2,$$

and the optimal convergence rate is achieved by estimator $\widehat{\boldsymbol{\Gamma}}$ given by (7) in Section 2.1, where C_* and C^* are positive constants free of n and p, $\| \cdot \|_d$ is a matrix norm with $d \ge 1$ in Section 2.2, $\mathcal{P}_q(\pi_n(p))$ denotes the minimax estimation problem for models (1)-(2) given in Theorem 3, constant $q \in [0, 1)$ and deterministic factor $\pi_n(p)$ characterize the sparsity of $\boldsymbol{\Gamma}$ specified by (8) and (11).

Because of the importance, large matrix estimation receives lots of attentions recently in statistics as well as in fields like remote sensing and finance. Classic optimal estimation procedures like sample covariance matrix estimator behave very poorly when the matrix size is comparable to or exceed the sample size (Bickel and Levina (2008a, b), Johnstone (2001) and Johnstone and Lu (2009)). In recent years various techniques have been developed for large covariance matrix estimation via sparsity. Wu and Pourahmadi (2003) explored the local stationary structure for nonparamertic estimation of large covariance matrices. Huang et al. (2006) used penalized likelihood method to estimate large covariance matrices. Yuan and Lin (2007) considered large covariance matrix estimation in Gaussian graph models. Bickel and Levina (2008a, b) developed regularization methods by banding or thresholding sample covariance matrix estimator or its inverse. Johnstone and Lu (2009) studied the consistent estimation of leading principal components when the matrix size is comparable to sample size. El Karoui (2008) employed the graph model approach to investigate sparsity structures and construct consistent estimators of large covariance matrices. Fan et. al. (2008) utilized factor models for estimating large covariance matrices. Lam and Fan (2009) studied sparsistency and convergence rates in large covariance matrix estimation. Cai et. al. (2010) and Cai and Zhou (2011) investigated minimax estimation of covariance matrices when both n and p are allowed to go to infinity and derived optimal convergence rates for estimating decaying or sparse covariance matrices. Wang and Zou (2010) considered the problem of estimating large volatility matrix based on noisy high-frequency financial data. Tao et. al. (2011) employed a matrix factor model to study the dynamics of large volatility matrices estimated from high-frequency financial data.

Models (1)-(2) capture the asymptotic essence of large volatility matrix estimation in high-frequency finance (Fan, Li and Yu (2011), Tao et. al. (2011), Wang and Zou (2010), Zheng and Li (2011)), where $X_i(t)$ correspond to true log prices, and ε_i are micro-structure noises in the observed high-frequency financial data. The models can also be viewed as a generalization of covariance matrix estimation in two scenarios. Take $\boldsymbol{\mu}_t = 0$ and $\boldsymbol{\sigma}_t$ to be constant in (1), then $\boldsymbol{\Gamma}$ is the covariance matrix of $\sqrt{n}\,[\mathbf{X}(t_\ell) - \mathbf{X}(t_{\ell-1})]$. For the case of no noise (i.e. $\varepsilon_i(t_\ell) = 0$), the problem becomes regular large covariance matrix estimation. For the noise case (i.e. $\varepsilon_i(t_\ell) \neq 0$), we may view models (1)-(2) as a complicated covariance matrix estimation problem where observed data are dependent and have measurement errors. The optimal convergence rate for sparse covariance matrix estimation obtained in Cai and Zhou (2011) is $\pi_n(p)\left(n^{-1/2}\sqrt{\log p}\right)^{1-q}$. Comparing it with the convergence rate in Theorem 1 we find that the dependence on sample size n is changed from $n^{-1/2}$ to $n^{-1/4}$. Such slower convergence rate is intrinsically due to the noise contamination in the observed data. When a univariate continuous diffusion process is observed with noise at n discrete time points, Gloter and Jacod (2001) showed that the optimal convergence rate for estimating univariate integrated volatility is $n^{-1/4}$. The convergence rate $n^{-1/4}$ in Theorem 1 matches with the optimal convergence rate for estimating univariate volatility based on noisy data. The phenomenon will be heuristically explained through model transformation in the derivation of the minimax lower bound.

Our approach to solving the minimax problem is as follows. We construct a multi-scale volatility matrix estimator and then threshold it to obtain threshold volatility matrix estimator. We show that the elements of the multi-scale volatility matrix estimator obey sub-Gaussian tail with optimal rate $n^{-1/4}$ and then demonstrate that the constructed estimator has convergence rate given by Theorem 1. As models (1)-(2) involve a nonparametric diffusion with measurement errors, it is generally very hard to derive the minimax risk for such a complicated problem. Fortunately we are able to find a clever way to establish the asymptotic optimality of the estimator. We first take $\boldsymbol{\mu}(t) = 0$ and $\boldsymbol{\sigma}(t)$ (thus $\boldsymbol{\Gamma}$) to be a constant matrix. The problem becomes covariance matrix estimation where the observed data are dependent and have measurement errors. Second, taking a special transformation we are able to convert the problem into a new covariance matrix estimation problem where observed data are independent but not identically distributed, with covariance matrices equal to $\boldsymbol{\Gamma}$ plus an identity matrix multiplying by a shrinking factor depending on sample size. Third adopting the minimax lower bound technique developed in Cai and Zhou (2011) for covariance matrix estimation based on i.i.d. data, we establish a minimax lower bound for estimating constant $\boldsymbol{\Gamma}$ based on the independent but non-identical observations. With the established minimax lower bound, we prove that the constructed estimator asymptotically achieves the minimax lower bound and thus is optimal.

The rest of the paper proceeds as follows. Section 2 presents the construction of volatility matrix estimator and establishes the asymptotic theory for the estimator as both n and p go to infinity. Section 3 derives the minimax lower bound for estimating large volatility matrix under models (1)-(2) and shows that the constructed estimator asymptotically achieves the minimax lower bound. Thus combining results in Sections 2 and 3 together we prove Theorem 1 in Section 1. To facilitate the reading we relegate all proofs to Sections 4 and 5, where we first provide main proofs of the theorems and then collect additional proofs of technical lemmas after the main proofs.

2 Volatility matrix estimation

2.1 Estimator

Let K be an integer and $\lfloor n/K \rfloor$ be the largest integer $\leq n/K$. We divide n time points $t_1, \cdots, t_n$ into K non-overlap groups $\boldsymbol{\tau}^k = \{t_\ell, \ell = k, K+k, 2K+k, \cdots\}$, $k = 1, \cdots, K$. 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 $k = 1, \cdots, K$, we write the r-th time point in $\boldsymbol{\tau}^k$ as $\tau_r^k = t_{(r-1)K+k}$, $r = 1, \cdots, |\boldsymbol{\tau}^k|$. With each $\boldsymbol{\tau}^k$, we define volatility matrix estimator

$$\tilde{\Gamma}_{ij}(\boldsymbol{\tau}^k) = \sum_{r=2}^{|\boldsymbol{\tau}^k|} [Y_i(\tau_r^k) - Y_i(\tau_{r-1}^k)][Y_j(\tau_r^k) - Y_j(\tau_{r-1}^k)], \quad \tilde{\boldsymbol{\Gamma}}(\boldsymbol{\tau}^k) = \left(\tilde{\Gamma}_{ij}(\boldsymbol{\tau}^k)\right)_{1\leq i,j\leq p}. \tag{3}$$

Here in (3) to account for noise 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

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

Let N be the largest integer $\leq n^{1/2}$, and $K_m = m + N$, $m = 1, \cdots, N$. We use each K_m to define a one-scale volatility matrix estimator $\tilde{\mathbf{\Gamma}}^{K_m}$ and then combine them together to form a multi-scale volatility matrix estimator

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

where

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

which satisfy

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

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

$$\widehat{\mathbf{\Gamma}} = \left(\tilde{\Gamma}_{ij} 1(|\tilde{\Gamma}_{ij}| \geq \varpi) \right), \tag{7}$$

where ϖ is a threshold value.

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

2.2 Asymptotic theory

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

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

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

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

and

$$\|\mathbf{A}\|_2^2 \leq \|\mathbf{A}\|_1 \, \|\mathbf{A}\|_\infty.$$

For symmetric $\mathbf{A}$, $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$, and $\|\mathbf{A}\|_2$ is equal to the largest absolute eigenvalue of $\mathbf{A}$.

Second we state some technical conditions for asymptotic analysis.

A1. Assume that noise $\varepsilon_i(t_\ell)$ and diffusion process $\mathbf{X}(t)$ in models (1)-(2) are independent; $\varepsilon_i(t_\ell)$, $i = 1, \cdots, p$, $\ell = 1, \cdots, n$, are independent normal with mean zero and $Var[\varepsilon_i(t_\ell)] = \eta_i \leq \kappa$ for some positive constant κ. Also assume $n^{\beta/2} \leq p \leq \exp(\beta_0\sqrt{n})$ for some constants $\beta > 1$ and $\beta_0 > 0$.

A2. Assume that each component of drift $\boldsymbol{\mu}(t)$ has bounded variation, and

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

where c_1 and c_2 are positive constants.

A3. We impose sparsity on $\mathbf{\Gamma}$,

$$\sum_{j=1}^{p}|\Gamma_{ij}|^q \leq \Psi\,\pi_n(p), \quad i = 1, \cdots, p, \tag{8}$$

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 $\log p$.

In Condition A1, the first part is a typical assumption in the literature of measurement error models; $p \geq n^{\beta/2}$ is required for obtaining the minimax lower bound in Theorem 1, 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 $\mathbf{\Gamma}$, otherwise the minimax risk in Theorem 1 will be bounded below from zero. Condition A2 is to impose proper assumptions on drift and volatility of the diffusion process so that we can obtain sub-Gaussian tail probability for volatility matrix estimator $\tilde{\mathbf{\Gamma}}$. Condition A3 is a common sparsity assumption required for consistently estimating large matrices (Bickel and Levina (2008b), Cai and Zhou (2011), and Johnstone and Lu (2009)).

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

Theorem 2 *Under Models (1)-(2) and Conditions A1-A2, estimator $\tilde{\mathbf{\Gamma}}$ in (5) satisfies that for $1 \leq i, j \leq p$ and positive x in a neighbor of* 0,

$$\mathbb{P}\left(\left|\tilde{\Gamma}_{ij} - \Gamma_{ij}\right| \geq x\right) \leq \varsigma_1 \exp\left\{\log(n/x) - \sqrt{n}\,x^2/\varsigma_0\right\}, \tag{9}$$

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

Remark 1. Theorem 2 establishes sub-Gaussian tail for the elements of matrix estimator $\tilde{\mathbf{\Gamma}}$. It is known that, when a univariate continuous diffusion process is observed with noise at n discrete time points, the optimal convergence rate for estimating integrated volatility is $n^{-1/4}$ (Gloter and Jacod (2001), Fan and Wang (2007) and Zhang (2006)). The $\sqrt{n}\,x^2$ factor in the exponent of the tail probability bound in (9) indicates the $n^{-1/4}$ convergence rate for $\tilde{\Gamma}_{ij} - \Gamma_{ij}$, which matches the optimal convergence rate for estimating univariate integrated volatility. This is in comparison with sub-optimal convergence results on limiting distributions and tail probabilities in the literature where the $n^{-1/6}$ convergence rate was obtained, see for example Fan et. al. (2011), Wang and Zou (2010)), and Zheng and Li (2011).

Theorem 3 *For threshold estimator* $\widehat{\mathbf{\Gamma}}$ *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 tail probability bound in (9). Denote by* $\mathcal{P}_q(\pi_n(p))$ *the set of distributions of* $Y_i(t_\ell)$*,* $i = 1, \cdots, p$*,* $\ell = 1, \cdots, n$*, from models (1)-(2) satisfying Conditions A1-A3. Then as* $n, p \to \infty$*,*

$$\sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\left\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\right\|_2^2 \leq \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\left\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\right\|_1^2 \leq C^* \left[\pi_n(p)\left(n^{-1/4}\sqrt{\log p}\right)^{1-q}\right]^2, \tag{10}$$

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

Remark 2. The convergence rate obtained in Theorem 3 increases with sample size n and matrix size p through $n^{-1/4}\sqrt{\log p}$. In section 3 we will establish the minimax lower bound for estimating $\mathbf{\Gamma}$ and show that the convergence rate in Theorem 3 is asymptotically optimal. For sparse covariance matrix estimation, Cai and Zhou (2011) has shown that the thresholding estimator in Bickel and Levina (2008b) is optimal and the optimal convergence rate depends on n and p through $n^{-1/2}\sqrt{\log p}$. In comparison of Theorem 3 with the optimal convergence rate for covariance matrix estimation, the convergence rate in Theorem 3 has a similar form but depends on n in terms of $n^{-1/4}$ instead of $n^{-1/2}$ for covariance matrix estimation. The slower convergence rate here is intrinsically due to the noise contamination in the observed data under our set-up. The $n^{-1/4}$ convergence rate conforms with the optimal convergence rate for estimating univariate integrated volatility, and will be heuristically explained in Remark 5 after the minimax lower bound result in Section 3.

3 Optimal convergence rate

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

Theorem 4 *For models (1)-(2) satisfying Conditions A1-A3, if for some* $M > 0$*,*

$$\pi_n(p) \leq M n^{(1-q)/4} / (\log p)^{(3-q)/2}, \tag{11}$$

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

$$\inf_{\check{\mathbf{\Gamma}}} \sup_{\mathcal{P}_q(\pi_n(p))} \mathbb{E}\left\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}\right\|_2^2 \geq C_* \left[\pi_n(p)\left(n^{-1/4}\sqrt{\log p}\right)^{1-q}\right]^2, \tag{12}$$

where C_* *is a positive constant.*

Remark 3. The lower bound convergence rate in Theorem 4 matches the convergence rate of estimator $\widehat{\boldsymbol{\Gamma}}$ obtained in Theorem 3. Since for a symmetric matrix $\mathbf{A}$, the Riesz-Thorin interpolation theorem implies that for all $d \geq 1$, $\|\mathbf{A}\|_d \leq \|\mathbf{A}\|_1 = \|\mathbf{A}\|_\infty$. Combining Theorems 3 and 4 together we prove Theorem 1 in Section 1, which shows that the optimal convergence rate is $\pi_n(p)\left(n^{-1/4}\sqrt{\log p}\right)^{1-q}$ and estimator $\widehat{\boldsymbol{\Gamma}}$ in (7) asymptotically achieves the optimal convergence rate.

Remark 4. Condition (11) is a technical condition that we need to establish the minimax lower bound. It is compatible with Assumptions A1 and A3 regarding the constraint on n and p and slow growth of $\pi_n(p)$ in sparsity condition (8).

We usually rely on LeCam's method or Assouad's lemma to establish minimax lower bound. For matrix estimation problem Cai and Zhou (2011) has developed an approach combining both LeCam's method and Assouad's lemma to deal with minimax lower bound in estimating covariance matrix based on i.i.d. observations. The observations from Models (1)-(2) are noisy and dependent. Luckily we are able to find a nice trick that takes a special subset of matrices and transforms the problem into a new covariance matrix estimation problem with independent but non-identical observations. We then adopt the approach in Cai and Zhou (2011) to derive the minimax lower bound for the new covariance matrix estimation problem with independent but non-identical data. Thus, we prove Theorem 4.

3.1 Model transformation

We take diffusion drift $\boldsymbol{\mu}_t = 0$ and diffusion matrix $\boldsymbol{\sigma}_t$ to be constant matrix $\boldsymbol{\sigma}$, then $\boldsymbol{\Gamma} = (\Gamma_{ij}) = \boldsymbol{\sigma}^T \boldsymbol{\sigma}$, and the sparsity condition (8) becomes

$$\sum_{j=1}^{p} |\Gamma_{ij}|^q \leq c\, \pi_n(p), \tag{13}$$

where $c = E(\Psi)$ and Ψ is given by (8).

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

$$\mathbf{Y}_l = \boldsymbol{\sigma}\mathbf{B}_{t_l} + \boldsymbol{\varepsilon}_l, \qquad l = 1, \cdots, n,\ t_l = l/n, \tag{14}$$

and we assume $\boldsymbol{\varepsilon}_l \sim N(0, \kappa^2 I_p)$, where $\kappa > 0$ is specified in Condition A1. As $\mathbf{Y}_l$ are dependent, we take differences in (14) and obtain

$$\mathbf{Y}_l - \mathbf{Y}_{l-1} = \boldsymbol{\sigma}(\mathbf{B}_{t_l} - \mathbf{B}_{t_{l-1}}) + \boldsymbol{\varepsilon}_l - \boldsymbol{\varepsilon}_{l-1}, \qquad l = 1, \cdots, n, \tag{15}$$

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 at different rows are independent but have correlation at the same row. At the i-th row, elements $\varepsilon_i(t_l) - \varepsilon_i(t_{l-1})$, $l = 1, \cdots, n$, have covariance matrix $\kappa^2\, \boldsymbol{\Upsilon}$, where $\boldsymbol{\Upsilon}$ is a $n \times n$ tridiagonal matrix with 2 for diagonal entries, -1 for next to diagonal entries, and 0 elsewhere. $\boldsymbol{\Upsilon}$ is a Toeplitz matrix (Wilkinson (1988)) that can be diagonalized as follows,

$$\boldsymbol{\Upsilon} = \mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^T, \qquad \boldsymbol{\Phi} = diag(\varphi_1, \cdots, \varphi_n), \tag{16}$$

where φ_l are eigenvalues with expressions

$$\varphi_l = 4\sin^2\left[\frac{\pi\, l}{2(n+1)}\right], \qquad l = 1,\cdots,n, \tag{17}$$

and $\mathbf{Q}$ is an orthogonal matrix formed by the eigenvectors of $\boldsymbol{\Upsilon}$. Using (16) we transform the i-th row of matrix $(\boldsymbol{\varepsilon}_l - \boldsymbol{\varepsilon}_{l-1}, 1 \le l \le n)$ by $\mathbf{Q}$, and obtain

$$Var[(\varepsilon_i(t_l) - \varepsilon_i(t_{l-1}), 1 \le l \le n)\,\mathbf{Q}] = \kappa^2 \mathbf{Q}^T \boldsymbol{\Upsilon} \mathbf{Q} = \kappa^2 \boldsymbol{\Phi},$$

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

$$\begin{aligned}(e_{il}, 1 \le l \le n) &= (\sqrt{n}[\varepsilon_i(t_l) - \varepsilon_i(t_{l-1})], 1 \le l \le n)\mathbf{Q},\\ (u_{il}, 1 \le l \le n) &= (\sqrt{n}[Y_i(t_l) - Y_i(t_{l-1})], 1 \le l \le n)\mathbf{Q},\\ (v_{il}, 1 \le l \le n) &= (\sqrt{n}[B_i(t_l) - B_i(t_{l-1})], 1 \le l \le n)\mathbf{Q}.\end{aligned}$$

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

Put (15) in a matrix form and 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 matrices (u_{il}), (v_{il}) and (e_{il}), respectively. Then the above matrix equation is equivalent to

$$\mathbf{U}_l = \boldsymbol{\sigma}\mathbf{V}_l + \mathbf{e}_l, \qquad l = 1,\cdots,n, \tag{18}$$

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 (18) we have that observed vectors $\mathbf{U}_1,\cdots,\mathbf{U}_n$ are independent with

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

where $a_l = 1 + \kappa^2\, n\, \varphi_l$ with $0 < \kappa < \infty$.

3.2 Lower bound

We convert the minimax lower bound problem stated in Theorem 4 into a much simpler problem of estimating $\boldsymbol{\Gamma}$ based on observations $\mathbf{U}_1,\cdots,\mathbf{U}_n$ from (18), where $\boldsymbol{\Gamma}$ are constant matrices satisfying (13) and $\|\boldsymbol{\Gamma}\|_2 \le \tau$ for some constant $\tau > 0$, and we denote the new minimax estimation problem by $\mathcal{Q}_q(\pi_n(p))$. The theorem below derives its minimax lower bound.

Theorem 5 *Assume $p \ge n^{\beta/2}$ for some $\beta > 1$. If $\pi_n(p)$ obeys (11), the minimax risk for estimating matrix $\boldsymbol{\Gamma}$ with $\mathcal{Q}_q(\pi_n(p))$ satisfies that as $n, p \to \infty$,*

$$\inf_{\check{\boldsymbol{\Gamma}}} \sup_{\mathcal{Q}_q(\pi_n(p))} \mathbb{E}\left\|\check{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right\|_2^2 \ge C_* \left[\pi_n(p)\left(n^{-1/4}\sqrt{\log p}\right)^{1-q}\right]^2, \tag{19}$$

where C_* is a positive constant.

Remark 5. As we discussed in Remarks 1 and 2 in Section 2.2, 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. Actually Section 3.1 reveals some intrinsical insight on the $n^{-1/4}$ convergence rate. The transformation in Section 3.1 converts model (15) with noisy data into model (18) where independent vector $\mathbf{U}_l$ follow multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Gamma} + \kappa^2 \, n \, \varphi_l \, \mathbf{I}_p$, $l = 1, \cdots, n$. The transformation is via orthogonal matrix $\mathbf{Q}$, which diagonalizes Toeplitz matrix $\mathbf{\Upsilon}$ and is equal to $(\sin(\ell r \pi/(n+1)), 1 \le \ell, r \le n)$ normalized by $\sqrt{2/(n+1)}$ (see Salkuyeh (2006)). Thus the transformation from model (15) to model (18) corresponds to a discrete sine transform, with (18) in frequency domain and $\mathbf{U}_l \sim N(0, \mathbf{\Gamma} + \kappa^2 \, n \, \varphi_l \, \mathbf{I}_p)$ corresponding to the discrete sine transform of the data at frequency $l\pi/(n+1)$. Note that (17) indicates that $n \, \varphi_l$ behave like l^2/n. Since for l with much higher order than $\sqrt{n}$, $\mathbf{\Gamma} + \kappa^2 \, n \, \varphi_l \, \mathbf{I}_p$ are dominated by $\kappa^2 \, n \, \varphi_l \, \mathbf{I}_p$, the corresponding $\mathbf{U}_l$ essentially behave like noise $\mathbf{e}_l$ and thus are not informative for estimating $\mathbf{\Gamma}$. On the other hand, for l up to the order of $\sqrt{n}$, $n \, \varphi_l$ are relatively small, and $\mathbf{\Gamma} + \kappa^2 \, n \, \varphi_l \, \mathbf{I}_p$ are close to $\mathbf{\Gamma}$, thus the corresponding $\mathbf{U}_l$ are statistically similar to $\mathbf{V}_l$. Hence, there are only $\sqrt{n}$ number of frequencies at which the transformed data $\mathbf{U}_l$ are informative for estimating $\mathbf{\Gamma}$, and we use these $\mathbf{U}_l$ to estimate covariance matrix $\mathbf{\Gamma}$ and obtain $(\sqrt{n})^{-1/2} = n^{-1/4}$ convergence rate. In fact, we have seen the phenomenon in Section 2.1 where scales used in the construction of $\tilde{\Gamma}$ in (5) correspond to the $\sqrt{n}$ number of K_m, which all are of order $\sqrt{n}$.

4 Proofs of Theorems 2 and 3

Denote by C's generic constants whose values are free of n and p and may change from appearance to appearance. 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 $C_1 \le u_{n,p}/v_{n,p} \le C_2$. Let

$$\mathbf{Y}_r^{k_m} = (Y_1(\tau_r^{k_m}), \cdots, Y_p(\tau_r^{k_m}))^T, \quad \mathbf{X}_r^{k_m} = (X_1(\tau_r^{k_m}), \cdots, X_p(\tau_r^{k_m}))^T,$$

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

the vectors corresponding to the data, diffusion process, and noise at time point $\tau_r^{k_m}$, $r = 1, \cdots, |\boldsymbol{\tau}^{k_m}|$, $k_m = 1, \cdots, K_m$, and $m = 1, \cdots, N$. Note that we choose index k_m to specify that the analyses are associated with the study of $\mathbf{\Gamma}^{K_m}$ here and below. We decompose $\tilde{\mathbf{\Gamma}}^{K_m}$

defined in (4) as follows,

$$
\begin{aligned}
\tilde{\mathbf{\Gamma}}^{K_m} &= \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left(\mathbf{Y}_r^{k_m}-\mathbf{Y}_{r-1}^{k_m}\right)\left(\mathbf{Y}_r^{k_m}-\mathbf{Y}_{r-1}^{k_m}\right)^T \\
&= \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left(\mathbf{X}_r^{k_m}-\mathbf{X}_{r-1}^{k_m}+\boldsymbol{\varepsilon}_r^{k_m}-\boldsymbol{\varepsilon}_{r-1}^{k_m}\right)\left(\mathbf{X}_r^{k_m}-\mathbf{X}_{r-1}^{k_m}+\boldsymbol{\varepsilon}_r^{k_m}-\boldsymbol{\varepsilon}_{r-1}^{k_m}\right)^T \\
&= \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left\{\left(\mathbf{X}_r^{k_m}-\mathbf{X}_{r-1}^{k_m}\right)\left(\mathbf{X}_r^{k_m}-\mathbf{X}_{r-1}^{k_m}\right)^T+\left(\boldsymbol{\varepsilon}_r^{k_m}-\boldsymbol{\varepsilon}_{r-1}^{k_m}\right)\left(\boldsymbol{\varepsilon}_r^{k_m}-\boldsymbol{\varepsilon}_{r-1}^{k_m}\right)^T\right. \\
&\left.+\left(\mathbf{X}_r^{k_m}-\mathbf{X}_{r-1}^{k_m}\right)\left(\boldsymbol{\varepsilon}_r^{k_m}-\boldsymbol{\varepsilon}_{r-1}^{k_m}\right)^T+\left(\boldsymbol{\varepsilon}_r^{k_m}-\boldsymbol{\varepsilon}_{r-1}^{k_m}\right)\left(\mathbf{X}_r^{k_m}-\mathbf{X}_{r-1}^{k_m}\right)^T\right\} \\
&\equiv \mathbf{V}^{K_m}+\mathbf{G}^{K_m}(1)+\mathbf{G}^{K_m}(2)+\mathbf{G}^{K_m}(3),
\end{aligned}
\tag{20}
$$

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

$$
\begin{aligned}
\tilde{\mathbf{\Gamma}} &= \sum_{m=1}^{N} a_m\tilde{\mathbf{\Gamma}}^{K_m}+\zeta(\tilde{\mathbf{\Gamma}}^{K_1}-\tilde{\mathbf{\Gamma}}^{K_N}) \\
&= \sum_{m=1}^{N} a_m\mathbf{V}^{K_m}+\zeta(\mathbf{V}^{K_1}-\mathbf{V}^{K_N})+\sum_{r=1}^{3}\left[\sum_{m=1}^{N} a_m\mathbf{G}^{K_m}(r)+\zeta(\mathbf{G}^{K_1}(r)-\mathbf{G}^{K_N}(r))\right] \\
&\equiv \mathbf{V}+\mathbf{G}(1)+\mathbf{G}(2)+\mathbf{G}(3),
\end{aligned}
\tag{21}
$$

where the terms denoted by $\mathbf{G}$'s are associated with noise, and the $\mathbf{V}$ term corresponds to process $\mathbf{X}_t$. We establish tail probabilities for these $\mathbf{V}$ and $\mathbf{G}$ terms in the following three propositions whose proofs will be given in the subsequent subsections.

Proposition 6 *Under the assumptions of Theorem 2, we have for $1 \le i, j \le p$ and positive d in a neighbor of 0,*

$$P\left(|V_{ij}-\Gamma_{ij}| \ge d\right) \le C_1 n \exp\left\{-\sqrt{n}d^2/C_2\right\}.$$

Proposition 7 *Under the assumptions of Theorem 2, we have for $1 \le i, j \le p$ and positive d in a neighbor of 0,*

$$P\left(|G_{ij}(1)| \ge d\right) \le C_1 \exp\left\{-\sqrt{n}d^2/C_2\right\}. \tag{22}$$

Proposition 8 *Under the assumptions of Theorem 2, we have for $1 \le i, j \le p$ and positive d in a neighbor of 0,*

$$
\begin{aligned}
P\left(|G_{ij}(2)| \ge d\right) &\le C_1(n^{1/2}/d)\exp\left\{-\sqrt{n}d^2/C_2\right\}, &(23)\\
P\left(|G_{ij}(3)| \ge d\right) &\le C_1(n^{1/2}/d)\exp\left\{-\sqrt{n}d^2/C_2\right\}. &(24)
\end{aligned}
$$

Proof of Theorem 2. From (21) we have

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

and thus the theorem is a consequence of Propositions 6-8. ■

Proof of Theorem 3. Define

$$\begin{aligned}A_{ij} &= \left\{|\widehat{\Gamma}_{ij}-\Gamma_{ij}|\leq 2\min\{|\Gamma_{ij}|,\varpi\}\right\},\\ D_{ij} &= (\widehat{\Gamma}_{ij}-\Gamma_{ij})1(A_{ij}^c),\quad \mathbf{D}=(D_{ij})_{1\leq i,j\leq p}.\end{aligned}$$

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

$$E||\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}||_2^2\leq E||\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}||_1^2\leq E||\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}-\mathbf{D}||_1^2+E||\mathbf{D}||_1^2. \tag{25}$$

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

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

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

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

see Lemma 1 in Wang and Zou (2010). The rest of the proof is to show that $E||\mathbf{D}||_1=O(n^{-2})$, a negligible term. Indeed, thresholding 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

$$\begin{aligned}&E||\mathbf{D}||_1^2=E\left[\max_{1\leq j\leq p}\sum_{i=1}^{p}|\widehat{\Gamma}_{ij}-\Gamma_{ij}|1(|\widehat{\Gamma}_{ij}-\Gamma_{ij}|>2\min\{|\Gamma_{ij}|,\varpi\})\right]^2\\ &\leq p\sum_{i,j=1}^{p}E\left[|\Gamma_{ij}|^2 1(|\Gamma_{ij}|>2\min\{|\Gamma_{ij}|,\varpi\})1(\widehat{\Gamma}_{ij}=0)\right]\\ &\qquad+p\sum_{i,j}^{p}E\left[|\tilde{\Gamma}_{ij}-\Gamma_{ij}|^2 1(|\tilde{\Gamma}_{ij}-\Gamma_{ij}|>2\min\{|\Gamma_{ij}|,\varpi\})1(\widehat{\Gamma}_{ij}=\tilde{\Gamma}_{ij})\right]\\ &\equiv I_1+I_2.\end{aligned}$$

For term I_1, we have

$$I_1 = p \sum_{i,j=1}^{p} E\left[|\Gamma_{ij}|^2 1(|\Gamma_{ij}| > 2\varpi) 1(|\tilde{\Gamma}_{ij}| < \varpi)\right]$$
$$\leq p \sum_{i,j=1}^{p} E\left[|\Gamma_{ij}|^2 1(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi)\right]$$
$$\leq Cp \sum_{i,j=1}^{p} P(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi)$$
$$\leq Cp^3 \exp\left\{\log(n/\varpi) - \sqrt{n}\varpi^2/\varsigma_0\right\} \leq Cn^{-2},$$

where the third inequality is from Theorem 2, 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,

$$I_2 = p \sum_{i,j=1}^{p} E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 1(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > 2\min\{|\Gamma_{ij}|, \varpi\}) 1(|\tilde{\Gamma}_{ij}| \geq \varpi)\right]$$
$$= p \sum_{i,j=1}^{p} E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 1(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > 2\min\{|\Gamma_{ij}|, \varpi\}) 1(|\Gamma_{ij}| \geq \varpi/2) 1(|\tilde{\Gamma}_{ij}| \geq \varpi)\right]$$
$$+ p \sum_{i,j=1}^{p} E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 1(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > 2\min\{|\Gamma_{ij}|, \varpi\}) 1(|\Gamma_{ij}| < \varpi/2) 1(|\tilde{\Gamma}_{ij}| \geq \varpi)\right]$$
$$\leq p \sum_{i,j=1}^{p} E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 1(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi)\right] + p \sum_{i,j=1}^{p} E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 1(|\Gamma_{ij}| < \varpi/2, |\tilde{\Gamma}_{ij}| \geq \varpi)\right]$$
$$\leq 2p \sum_{i,j=1}^{p} E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^2 1(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi/2)\right]$$
$$\leq 2p \sum_{i,j=1}^{p} \left\{E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^4\right] P\left(|\tilde{\Gamma}_{ij} - \Gamma_{ij}| > \varpi/2\right)\right\}^{1/2}$$
$$\leq Cp^3 \exp\left\{\log(n/\varpi)/2 - \sqrt{n}\varpi^2/(8\varsigma_0)\right\} \leq Cn^{-2},$$

where the third inequality is due to Hölder's inequality, the fourth inequality is from Theorem 2 and (26) below, and the last inequality is due to the fact that $\varpi = \hbar n^{-1/4}\sqrt{\log(np)}$ with $\hbar^2/(8\varsigma_0) > 3$.

$$\max_{1 \leq i,j \leq p} E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^4\right] \leq C. \tag{26}$$

To complete the proof we need to show (26). Let

$$\tilde{\boldsymbol{\eta}} = diag(\tilde{\eta}_1, \cdots, \tilde{\eta}_p), \qquad \tilde{\eta}_i = \frac{1}{2n} \sum_{\ell=2}^{n} [Y_i(t_\ell) - Y_i(t_{\ell-1})]^2. \tag{27}$$

Then

$$\tilde{\boldsymbol{\Gamma}}^{*K_m} = \tilde{\boldsymbol{\Gamma}}^{K_m} - 2\frac{n - K_m + 1}{K_m}\tilde{\boldsymbol{\eta}} \tag{28}$$

are called average realized volatility matrix (ARVM) estimators in Wang and Zou (2010). Applying Theorem 1 of Wang and Zou (2010) to $\tilde{\boldsymbol{\Gamma}}^{*K_m}$ we have for $1 \le i, j \le p$ and $1 \le m \le N$,

$$E(|\tilde{\Gamma}_{ij}^{*K_m} - \Gamma_{ij}|^4) \le C\ \left[(K_m n^{-1/2})^{-4} + K_m^{-2} + (n/K_m)^{-2} + K_m^{-4} + n^{-2}\right] \le C. \tag{29}$$

From (5), (6) and (28) and with simple algebraic manipulations we can express $\tilde{\boldsymbol{\Gamma}}$ by $\tilde{\boldsymbol{\Gamma}}^{*K_m}$ as follows,

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

and thus

$$\tilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma} = \sum_{m=1}^{N} a_m \left(\tilde{\boldsymbol{\Gamma}}^{*K_m} - \boldsymbol{\Gamma}\right) + \zeta\left[(\tilde{\boldsymbol{\Gamma}}^{*K_1} - \boldsymbol{\Gamma}) - (\tilde{\boldsymbol{\Gamma}}^{*K_N} - \boldsymbol{\Gamma})\right]. \tag{30}$$

Combining (29) and (30) and using (6) we conclude for $1 \le i, j \le p$,

$$E\left[|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^4\right] \le (N+2)^3 \left[\sum_{m=1}^{N} a_m^4 E(|\tilde{\Gamma}_{ij}^{*K_m} - \Gamma_{ij}|^4) + \zeta^4 E(|\tilde{\boldsymbol{\Gamma}}^{*K_1} - \Gamma_{ij}|^4 + |\tilde{\boldsymbol{\Gamma}}^{*K_N} - \Gamma_{ij}|^4)\right] \le C.$$

■

4.1 Proof of Proposition 6

From the definition of $V_{ij}^{K_m}$ in (20), we have

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

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

Note that $A = \sum_{m=1}^{N}|a_m| + 2\zeta \sim 9/2$. From the expression of $V_{ij}^{K_m} - \Gamma_{ij}$ we obtain that for $1 \le m \le N$,

$$\begin{aligned}
&P\left(\left|V_{ij}^{K_m} - \Gamma_{ij}\right| \ge d/A\right) \le P\left(\frac{1}{K_m}\sum_{k_m=1}^{K_m}\left|[X_i, X_j]^{(k_m)} - \int_0^1 \gamma_{ij}(s)ds\right| \ge d/A\right) \\
&\le \sum_{k_m=1}^{K_m} P\left(\left|[X_i, X_j]^{(k_m)} - \int_0^1 \gamma_{ij}(s)ds\right| \ge d/A\right) \\
&\le C_1\sqrt{n}\exp\left\{-\sqrt{n}d^2/C_2\right\},
\end{aligned}$$

where the last inequality is due to Lemma 9 below and the fact that the maximum distance between consecutive grids in τ^{k_m} is bounded by $K_m/n \leq 2/\sqrt{n}$. Therefore,

$$P\left(|V_{ij} - \Gamma_{ij}| \geq d\right) \leq P\left(\sum_{m=1}^{N} |a_m| \left|V_{ij}^{K_m} - \Gamma_{ij}\right| + \zeta\left(\left|V_{ij}^{K_1} - \Gamma_{ij}\right| + \left|V_{ij}^{K_N} - \Gamma_{ij}\right|\right) \geq d\right)$$
$$\leq \sum_{m=1}^{N} P\left(\left|V_{ij}^{K_m} - \Gamma_{ij}\right| \geq d/A\right) + P\left(\left|V_{ij}^{K_1} - \Gamma_{ij}\right| \geq d/A\right) + P\left(\left|V_{ij}^{K_N} - \Gamma_{ij}\right| \geq d/A\right)$$
$$\leq C_1 n \exp\left\{-\sqrt{n} d^2/C_2\right\}.$$

Lemma 9 *Under Model (1) and Condition A2, for any sequence* $0 = \nu_0 \leq \nu_1 < \nu_2 < \cdots < \nu_m \leq \nu_{m+1} = 1$ *satisfying* $\max_{1\leq r\leq m+1} |\nu_r - \nu_{r-1}| \leq C/m$*, we have for* $1 \leq i, j \leq p$ *and small* $d > 0$*,*

$$P\left(\left|\sum_{r=2}^{m}(X_i(\nu_r) - X_i(\nu_{r-1}))(X_j(\nu_r) - X_j(\nu_{r-1})) - \int_0^1 \gamma_{ij}(s)ds\right| \geq d\right)$$
$$\leq C_1 \exp\left(-md^2/C_2\right).$$

Proof. The same arguments in the proof of Lemma 3 in Fan et. al. (2011) lead to

$$P\left(\left|\sum_{r=2}^{m}(X_i(\nu_r) - X_i(\nu_{r-1}))(X_j(\nu_r) - X_j(\nu_{r-1})) - \int_{\nu_1}^{\nu_m} \gamma_{ij}(s)ds\right| \geq d\right) \leq C_1 \exp\left\{-md^2/C_2\right\}. \tag{31}$$

From Condition A2 we have

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

and then for $d > C/m$,

$$P\left(\left|\sum_{r=2}^{m}(X_i(\nu_r) - X_i(\nu_{r-1}))(X_j(\nu_r) - X_j(\nu_{r-1})) - \int_0^1 \gamma_{ij}(s)ds\right| \geq d\right)$$
$$\leq P\left(\left|\sum_{r=2}^{m}(X_i(\nu_r) - X_i(\nu_{r-1}))(X_j(\nu_r) - X_j(\nu_{r-1})) - \int_{\nu_1}^{\nu_m} \gamma_{ij}(s)ds\right| \geq d - C/m\right)$$
$$\leq C_1 \exp\left\{-m(d - C/m)^2/C_2\right\} \leq C_1 \exp\left\{-md^2/C_2\right\},$$

where the second inequality is from (31) and the last inequality is due to the fact that $-m(d - C/m)^2 = -md^2 + 2C - C/m \leq -md^2 + 2C$. This proves the lemma for $d > C/m$.

For $d \leq C/m$, the tail probability bound in the lemma

$$C_1 \exp\left\{-md^2/C_2\right\} \geq C_1 \exp\left\{-C^2/(mC_2)\right\} \geq C_1 \exp\left\{-C^2/C_2\right\},$$

and we easily show the probability inequality in the lemma by choosing C_1 and C_2 so that $C_1 \exp\{-C^2/C_2\} \geq 1$. The proof is completed. ∎

4.2 Proof of Proposition 7

From the definition of $\mathbf{G}(1) = (G_{ij}(1))$ in (21), we obtain that $P(|G_{ij}(1)| \geq d)$ is bounded by

$$P\left(\left|\sum_{m=1}^{N} a_m G_{ij}^{K_m}(1) - 2\eta_i 1(i=j)\right| \geq d/2\right) + P\left(\left|\zeta(G_{ij}^{K_1}(1) - G_{ij}^{K_N}(1)) - 2\eta_i 1(i=j)\right| \geq d/2\right)$$
$$\leq C_1 \exp\{-\sqrt{n}d^2/C_2\} + C_3 \exp\{-nd^2/C_4\} \leq C_1 \exp\{-\sqrt{n}d^2/C_2\},$$

where the first inequality is from Lemmas 10 and 11 below.

Lemma 10 *Under the assumptions of Theorem 2, we have for* $1 \leq i, j \leq p$,

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

Proof. From the definition of $\mathbf{G}^{K_m} = (G_{ij}^{K_m}(1))$ in (20), we have

$$G_{ij}^{K_m}(1) = \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|} \left[\varepsilon_i(\tau_r^{k_m}) - \varepsilon_i(\tau_{r-1}^{k_m})\right] \left[\varepsilon_j(\tau_r^{k_m}) - \varepsilon_j(\tau_{r-1}^{k_m})\right]$$
$$= \frac{1}{K_m} \sum_{k_m=1}^{K_m} \sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|} \left[\varepsilon_i(\tau_r^{k_m})\varepsilon_j(\tau_r^{k_m}) - \varepsilon_i(\tau_r^{k_m})\varepsilon_j(\tau_{r-1}^{k_m}) - \varepsilon_i(\tau_{r-1}^{k_m})\varepsilon_j(\tau_r^{k_m}) + \varepsilon_i(\tau_{r-1}^{k_m})\varepsilon_j(\tau_{r-1}^{k_m})\right]$$
$$= \frac{2}{K_m} \sum_{r=1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_r) - \left[\frac{1}{K_m} \sum_{r=K_m+1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_{r-K_m}) + \frac{1}{K_m} \sum_{r=K_m+1}^{n} \varepsilon_i(t_{r-K_m})\varepsilon_j(t_r)\right]$$
$$- \left[\frac{1}{K_m} \sum_{r=1}^{K_m} \varepsilon_i(t_r)\varepsilon_j(t_r) + \frac{1}{K_m} \sum_{r=n-K_m+1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_r)\right]$$
$$\equiv I_1^{K_m} - I_2^{K_m} - I_3^{K_m},$$

and

$$\sum_{m=1}^{N} a_m G_{ij}^{K_m}(1) = \sum_{m=1}^{N} a_m I_1^{K_m} - \sum_{m=1}^{N} a_m I_2^{K_m} - \sum_{m=1}^{N} a_m I_3^{K_m}. \tag{32}$$

First of all, the fact that $\sum_{m=1}^{N} a_m/K_m = 0$ implies

$$\sum_{m=1}^{N} a_m I_1^{K_m} = \sum_{m=1}^{N} \frac{a_m}{K_m} \sum_{r=1}^{n} \varepsilon_i(t_r)\varepsilon_j(t_r) = 0. \tag{33}$$

Second, $I_2^{K_m}$ can be written as a quadratic form with matrix expression

$$\sum_{m=1}^{N} a_m\, I_2^{K_m} \equiv \boldsymbol{\varepsilon}_i^T \boldsymbol{\Lambda} \boldsymbol{\varepsilon}_j = \sum_{r=1}^{n} \sqrt{\eta_i \eta_j} \lambda_r Z_i(t_r) Z_j(t_r),$$

where $\mathbf{\Lambda} = (\Lambda_{r\ell}, r, \ell = 1, \cdots, n)$ is a n by n symmetric matrix with $\Lambda_{rr} = 0$, $\Lambda_{r,r\pm K_m} = a_m/K_m$ for $m = 1, \cdots, N$, and zero otherwise, $\{\lambda_r, r = 1, \cdots, n\}$ are the eigenvalues of $\mathbf{\Lambda}$, $\boldsymbol{\varepsilon}_i = (\varepsilon_i(t_1), \varepsilon_i(t_2), \cdots, \varepsilon_i(t_n))^T$, and $Z_i(t_r), r = 1, \cdots, n$, $i = 1, \cdots, p$, are i.i.d. standard normal random variables. Simple calculations show

$$\max_{1\le r\le n} |\lambda_r| = ||\mathbf{\Lambda}||_2 \le ||\mathbf{\Lambda}||_1 \le 2\sum_{m=1}^{N} \frac{|a_m|}{|K_m|} \asymp 1/N, \quad \sum_{r=1}^{n} \lambda_r = tr(\mathbf{\Lambda}) = 0.$$

Both $Z_i(t_r)Z_j(t_r), i \ne j$ and $Z_i(t_r)^2 - 1$ satisfy Condition (P) and Equation (3.12) on page 45 of Saulis and Statulevičius (1991), and

$$E\left(\sum_{m=1}^{N} a_m\, I_2^{K_m}\right) = 0,$$

$$Var\left(\sum_{m=1}^{N} a_m\, I_2^{K_m}\right) = 2[1 + 1(i=j)]\eta_i\eta_j \sum_{m=1}^{N}\sum_{r=K_m+1}^{n}\left(\frac{a_m}{K_m}\right)^2 \asymp \frac{n}{N^3},$$

thus applying Theorem 3.2 on page 45 of Saulis and Statulevičius (1991), we obtain

$$P\left(\left|\sum_{m=1}^{N} a_m I_2^{K_m}\right| \ge d\right) \le C_1 \exp\left\{-(N^3/n)d^2/C_2\right\} = C_1 \exp\{-\sqrt{n}d^2/C_2\}. \tag{34}$$

Finally we will prove the result for $\sum_{m=1}^{N} a_m I_3^{K_m}$. Due to the similarity we show the result only for its first term.

$$\begin{aligned}
&\sum_{m=1}^{N} \frac{a_m}{K_m} \sum_{r=1}^{K_m} \varepsilon_i(t_r)\varepsilon_j(t_r) = \sum_{m=1}^{N} \frac{a_m}{K_m} \sum_{r=1}^{K_1} \varepsilon_i(t_r)\varepsilon_j(t_r) + \sum_{m=2}^{N} \frac{a_m}{K_m} \sum_{r=K_1+1}^{K_m} \varepsilon_i(t_r)\varepsilon_j(t_r) \\
&= \left(\sum_{m=1}^{N} \frac{a_m}{K_m}\right)\left(\sum_{r=1}^{K_1} \varepsilon_i(t_r)\varepsilon_j(t_r)\right) + \sum_{r=K_1+1}^{K_N}\left(\sum_{m=r-K_1+1}^{N} \frac{a_m}{K_m}\right)\varepsilon_i(t_r)\varepsilon_j(t_r) \\
&= \sum_{r=K_1+1}^{K_N} R_r \varepsilon_i(t_r)\varepsilon_j(t_r),
\end{aligned}$$

where $R_r = \left(\sum_{m=r-K_1+1}^{N} a_m/K_m\right)$. Simple algebraic manipulations get $\max_r |R_r| \sim 1/N$, and

$$E\left(\sum_{r=K_1+1}^{K_N} R_r\varepsilon_i(t_r)\varepsilon_j(t_r)\right) = \eta_i \cdot 1(i=j),$$

$$Var\left(\sum_{r=K_1+1}^{K_N} R_r\varepsilon_i(t_r)\varepsilon_j(t_r)\right) = [1 + 1(i=j)]\eta_i\eta_j \sum_{r=K_1+1}^{K_N}\left(\sum_{m=r-K_1+1}^{N} \frac{a_m}{K_m}\right)^2 \asymp \frac{1}{N}.$$

An application of Theorem 3.2 on page 45 of Saulis and Statulevičius (1991) leads to

$$P\left(\left|\sum_{m=1}^{N} a_m I_3^{K_m} - 2\eta_i \cdot 1(i=j)\right| \geq d\right) \leq C_1 \exp\left\{-\sqrt{n}d^2/C_2\right\}. \tag{35}$$

Combining (32)-(35) we conclude

$$\begin{aligned}
&P\left(|G_{ij}(1) - 2\eta_i \cdot 1(i=j)| \geq d\right) \\
&\leq P\left(\left|\sum_{m=1}^{N} a_m I_2^{K_m}\right| \geq d/2\right) + P\left(\left|\sum_{m=1}^{N} a_m I_3^{K_m} - 2\eta_i \cdot 1(i=j)\right| \geq d/2\right) \\
&\leq C_1 \exp\left\{-\sqrt{n}d^2/C_2\right\}.
\end{aligned}$$

■

Lemma 11 *Under the assumptions of Theorem 2, we have for* $1 \leq i, j \leq p$,

$$P\left(\left|\zeta(G_{ij}^{K_1}(1) - G_{ij}^{K_N}(1)) - 2\eta_i 1(i=j)\right| \geq d\right) \leq C_1 \exp\{-nd^2/C_2\}. \tag{36}$$

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

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

For $i \neq j$, using Lemma A.3 in Bickel and Levina (2008a) and $K_N, N \sim \sqrt{n}$, we have

$$P(|R_k| \geq d) \leq C_1 \exp\{-nd^2/C_2\}, \qquad 1 \leq k \leq 4. \tag{37}$$

For $i = j$, due to similarity, we will provide arguments only for R_1 and R_2. For R_1, with $i = j$, $\varepsilon_i^2(t_r) - \eta_i = \eta_i(V_r^2 - 1)$, where V_r^2 are i.i.d. χ_1^2 random variables, which satisfy Condition (P) and Equation (3.12) on page 45 of Saulis and Statulevičius (1991). Therefore,

$$P(|R_1| \geq d) = P\left(\left|\frac{K_N}{n(N-1)}\eta_i \sum_{r=K_1+1}^{n} (V_r^2 - 1)\right| \geq d\right) \leq C_1 \exp\{-nd^2/C_2\}. \tag{38}$$

Regarding to R_2 we have

$$\begin{aligned}
|R_2| &= \frac{K_N}{n(N-1)} \left|\sum_{r=K_1+1}^{n} \varepsilon_i(t_r)\varepsilon_i(t_{r-K_1})\right| = \frac{K_N}{n(N-1)} \left|\sum_{k_1=1}^{K_1} \sum_{r=2}^{|\boldsymbol{\tau}^{k_1}|} \varepsilon_i(\tau_r^{k_1})\varepsilon_i(\tau_{r-1}^{k_1})\right| \\
&\leq \left|\frac{K_N}{n(N-1)} \sum_{k_1=1}^{K_1} \sum_{r=1}^{\lfloor \frac{|\boldsymbol{\tau}^{k_1}|}{2} \rfloor} \varepsilon_i(\tau_{2r}^{k_1})\varepsilon_i(\tau_{2r-1}^{k_1})\right| + \left|\frac{K_N}{n(N-1)} \sum_{k_1=1}^{K_1} \sum_{r=1}^{\lfloor \frac{|\boldsymbol{\tau}^{k_1}|-1}{2} \rfloor} \varepsilon_i(\tau_{2r+1}^{k_1})\varepsilon_i(\tau_{2r}^{k_1})\right| = R_2^1 + R_2^2,
\end{aligned}$$

where $\varepsilon_i(\cdot)$'s in R_2^1 and R_2^2 are independent. Lemma A.3 in Bickel and Levina (2008a) infers that for $r = 1, 2$, $P(|R_2^r| \geq d) \leq C_1 \exp\{-nd^2/C_2\}$. Hence, we have

$$P(|R_2| \geq d) \leq P(|R_2^1| \geq d/2) + P(|R_2^2| \geq d/2) \leq C_1 \exp\{-nd^2/C_2\}. \tag{39}$$

Finally (37)-(39) together conclude

$$P\left(\left|\zeta G_{ij}^{K_1}(1) - \frac{K_N(n-K_1)}{n(N-1)}\eta_i 1(i=j)\right| \geq d\right) \leq C_1 \exp\{-nd^2/C_2\}.$$

We can establish a similar tail probability result for $\zeta G_{ij}^{K_N}(1)$ term and thus prove the lemma. ■

4.3 Proof of Proposition 8

As the proofs for $G_{ij}(2)$ and $G_{ij}(3)$ are similar, we give arugments only for $G_{ij}(2)$. As Lemma 12 below gives the tail probability for $G_{ij}^{K_m}(2)$ we have

$$P\left(|G_{ij}(2)| \geq d\right) \leq \sum_{m=1}^{N} P\left(\left|G_{ij}^{K_m}(2)\right| \geq d/A\right) + P\left(\left|G_{ij}^{K_1}(2)\right| \geq d/A\right) + P\left(\left|G_{ij}^{K_N}(2)\right| \geq d/A\right)$$
$$\leq C_1(n^{1/2}/d)\exp\left\{-\sqrt{n}d^2/C_2\right\},$$

where $A = \sum_{m=1}^{N} |a_m| + 2\zeta \sim 9/2$.

Lemma 12 *Under the assumptions of Theorem 2, we have for* $1 \leq i, j \leq p$,

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

Proof. Simple algebra shows

$$\begin{aligned}
G_{ij}^{K_m}(2) &= \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left[X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})\right]\left[\varepsilon_j(\tau_r^{k_m}) - \varepsilon_j(\tau_{r-1}^{k_m})\right] \\
&= \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left[X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})\right]\varepsilon_j(\tau_r^{k_m}) - \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left[X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})\right]\varepsilon_j(\tau_{r-1}^{k_m}) \\
&\equiv R_5^{K_m} + R_6^{K_m}.
\end{aligned}$$

Due to similarity, we prove only the result for $R_5^{K_m}$. Conditional on the whole path of $\mathbf{X}$, $R_5^{K_m}$ is the weighted sum of independent noise $\varepsilon_j(\cdot)$. Hence,

$$
\begin{aligned}
&P(|R_5^{K_m}| \geq d) = E\left[P(|R_5^{K_m}| \geq d|\mathbf{X})\right] \\
&= 2E\left[P\left(\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left[X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})\right]\varepsilon_j(\tau_r^{k_m}) \geq dK_m \middle| \mathbf{X}\right)\right] \\
&\leq E\left[C\frac{\sqrt{V_{ii}^{K_m}\eta_j}}{d\sqrt{K_m}}\exp\left\{-\frac{d^2K_m}{2V_{ii}^{K_m}\eta_j}\right\}\right] \\
&= E\left[C\frac{\sqrt{V_{ii}^{K_m}\eta_j}}{d\sqrt{K_m}}\exp\left\{-\frac{d^2K_m}{2V_{ii}^{K_m}\eta_j}\right\}1(\Omega)\right] + E\left[C\frac{\sqrt{V_{ii}^{K_m}\eta_j}}{d\sqrt{K_m}}\exp\left\{-\frac{d^2K_m}{2V_{ii}^{K_m}\eta_j}\right\}1(\Omega^c)\right] \\
&\equiv R_{5,1}^{K_m} + R_{5,2}^{K_m},
\end{aligned}
$$

where the inequality is due to the facts that $R_5^{K_m}|\mathbf{X} \sim N(0, V_{ii}^{K_m}\eta_j K_m)$, and for $Z \sim N(0,1)$, $P(Z \geq z) \leq C\exp\{-z^2/2\}/z$, and $V_{ii}^{K_m}$ is defined in (20), i.e.

$$
V_{ii}^{K_m} = \frac{1}{K_m}\sum_{k_m=1}^{K_m}[X_i, X_i]^{k_m} = \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{r=2}^{|\boldsymbol{\tau}^{k_m}|}\left[X_i(\tau_r^{k_m}) - X_i(\tau_{r-1}^{k_m})\right]^2, \tag{40}
$$

$$
\Omega = \left\{|V_{ii}^{K_m} - \Gamma_{ii}| \geq d\right\}. \tag{41}
$$

From the definition of Ω and for small d we obtain

$$
R_{5,2}^{K_m} = E\left[C\frac{\sqrt{V_{ii}^{K_m}\eta_j}}{d\sqrt{K_m}}\exp\left\{-\frac{d^2K_m}{2V_{ii}^{K_m}\eta_j}\right\}1(\Omega^c)\right] \leq \frac{C_1}{d\sqrt{K_m}}\exp\left\{-\frac{K_m d^2}{C_2}\right\}.
$$

On the other hand,

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

where the second inequality is due to Hölder inequality, and the last inequality is from Lemma 9 and the proof in Proposition 6 that

$$
P(\Omega) \leq C_1\sqrt{n}\exp\left\{-\sqrt{n}d^2/C_2\right\}.
$$

Therefore,

$$
\begin{aligned}
&P(|G_{ij}(2)^{K_m}| \geq d/(4A)) \leq \frac{C_1}{d\sqrt{K_m}}\exp\left\{-\frac{K_m d^2}{C_2}\right\} + (C_1/d)\exp\left\{-\frac{\sqrt{n}d^2}{C_2}\right\} \\
&\leq (C_1/d)\exp\left\{-\sqrt{n}d^2/C_2\right\}.
\end{aligned}
$$

■

5 Proofs of Theorems 4 and 5

Section 3.1 shows that Theorem 4 is a consequence of Theorem 5. The proof of Theorem 5 is similar to but more involved than the proof of Theorem 2 in Cai and Zhou (2011) 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 $\mathcal{G}_q(\pi_n(p))$ 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 (13) and $\|\mathbf{\Gamma}\|_2 \le \tau$ for constant $\tau > 0$. The second step applies the lower bound argument in Cai and Zhou (2011, Lemma 3) to the carefully constructed parameter set $\mathcal{F}_*$. In the third step we calculate the factor α defined in (50) below and the total variation affinity between two average of products of n independent but not identically 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 $b = (v_j)_{1 \le j \le p}$ such that $v_j = 0$ for $1 \le j \le p - r$ and $v_j = 0$ or 1 for $p - r + 1 \le j \le p$ under the constraint $\|b\|_0 = k$ (to be specified later). Each element $\lambda = (b_1, ..., b_r) \in B^r$ 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 $1 \le m \le r$, 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 $\lambda = (\lambda_1, ..., \lambda_r) \in \Lambda$ 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 $\gamma = (\gamma_1, ..., \gamma_r) \in \Delta$ and $\lambda = (\lambda_1, ..., \lambda_r) \in \Lambda$, we associate $\theta = (\gamma_1, ..., \gamma_r, \lambda_1, ..., \lambda_r)$ with a volatility matrix $\mathbf{\Gamma}(\theta)$ by

$$\mathbf{\Gamma}(\theta) = \mathbf{I}_p + \epsilon_{n,p} \sum_{m=1}^{r} \gamma_m A_m(\lambda_m). \tag{42}$$

For simplicity we assume that $\tau > 1$ in the assumption (13), otherwise we replace $\mathbf{I}_p$ in (42) by $c\mathbf{I}_p$ with a small constant $c > 0$. Finally we define $\mathcal{F}_*$ to be a collection of covariance matrices as

$$\mathcal{F}_* = \left\{ \mathbf{\Gamma}(\theta) : \mathbf{\Gamma}(\theta) = \mathbf{I}_p + \epsilon_{n,p} \sum_{m=1}^{r} \gamma_m A_m(\lambda_m),\ \theta = (\gamma, \lambda) \in \Theta \right\}. \tag{43}$$

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:

$$\epsilon_{n,p} = \upsilon \left(\frac{\log p}{\sqrt{n}} \right)^{1/2}, \qquad k = \left\lceil \frac{1}{2} \pi_n(p) \epsilon_{n,p}^{-q} \right\rceil - 1, \tag{44}$$

where υ is a fixed small constant that we require

$$0 < \upsilon < \left[\min\left\{\frac{1}{3}, \tau - 1\right\}\frac{1}{M}\right]^{\frac{1}{1-q}} \tag{45}$$

and

$$0 < \upsilon^2 < \frac{\beta - 1}{27 c_\kappa \beta}, \tag{46}$$

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

$$\sum_{l=1}^{n} a_l^{-2} \leq c_\kappa \sqrt{n}, \tag{47}$$

since

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

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

$$2k\epsilon_{n,p} \leq \pi_n(p)\epsilon_{n,p}^{1-q} \leq M\upsilon^{1-q} < \min\left\{\frac{1}{3}, \tau - 1\right\}, \tag{48}$$

and consequently every $\mathbf{\Gamma}(\theta)$ is diagonally dominant and positive definite, and $\|\mathbf{\Gamma}(\theta)\|_2 \leq \|\mathbf{\Gamma}(\theta)\|_1 \leq 2k\epsilon_{n,p} + 1 < \tau$. 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

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

where $\theta \in \Theta$, and we denote the joint distribution by P_θ. Applying Lemma 3 in Cai and Zhou (2011) to the parameter space Θ, we have

$$\inf_{\check{\mathbf{\Gamma}}}\max_{\theta\in\Theta} 2^2 \mathbb{E}_\theta \left\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}(\theta)\right\|_2^2 \geq \alpha \cdot \frac{r}{2} \cdot \min_{1\leq i\leq r}\left\|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\right\|, \tag{49}$$

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

$$\alpha \equiv \min_{\{(\theta,\theta'):H(\gamma(\theta),\gamma(\theta'))\geq 1\}} \frac{\|\mathbf{\Gamma}(\theta) - \mathbf{\Gamma}(\theta')\|_2^2}{H(\gamma(\theta),\gamma(\theta'))}, \qquad H(\gamma(\theta),\gamma(\theta')) = \sum_{i=1}^{r}|\gamma_i(\theta) - \gamma_i(\theta')|, \tag{50}$$

and

$$\bar{\mathbb{P}}_{i,a} = \frac{1}{2^{r-1}D_\Lambda}\sum_{\theta\in\Theta}\mathbb{P}_\theta \cdot \{\theta : \gamma_i(\theta) = a\}, \tag{51}$$

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 $\min_i \left\|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\right\|$ in (49). A lower bound for α is given by the following lemma whose proof is the same as that of Lemma 5 in Cai and Zhou (2011).

Lemma 13 *For α defined in Equation (50) we have*

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

A lower bound for $\min_i \left\|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\right\|$ is provided by the lemma below. Since its proof is long and very much involved, the proof details are collected in Section 5.1.

Lemma 14 *Let $\mathbf{U}_j$ be independent with $\mathbf{U}_j \sim N\left(0, \mathbf{\Gamma} + (a_l - 1)\,\mathbf{I}_p\right)$ 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 (51). Then there exists a constant $c_1 > 0$ such that*

$$\min_{1\leq i\leq r} \left\|\bar{\mathbb{P}}_{i,0} \wedge \bar{\mathbb{P}}_{i,1}\right\| \geq c_1$$

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 (49) and the bounds in Lemmas 13 and 14,

$$\begin{aligned}
\inf_{\check{\mathbf{\Gamma}}} \sup_{\mathcal{G}_q(\pi_n(p))} \mathbb{E}\left\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}\right\|_2^2 &\geq \max_{\mathbf{\Gamma}(\theta)\in\mathcal{F}_*} \mathbb{E}_\theta \left\|\check{\mathbf{\Gamma}} - \mathbf{\Gamma}\left(\theta\right)\right\|_2^2 \geq \frac{\left(k\epsilon_{n,p}\right)^2}{p} \cdot \frac{r}{8} \cdot c_1 \\
&\geq \frac{c_1}{16}(k\epsilon_{n,p})^2 = c_2 \pi_n^2(p) \left(n^{-1/4}\sqrt{\log p}\right)^{2-2q},
\end{aligned}$$

for some constant $c_2 > 0$. ∎

5.1 Proof of Lemma 14

We break the proof into a few major technical lemmas which are proved in Sections 5.3-5.4. Without loss of generality we consider only the case $i = 1$ and prove that there exists a constant $c_1 > 0$ such that $\left\|\bar{\mathbb{P}}_{1,0} \wedge \bar{\mathbb{P}}_{1,1}\right\| \geq c_1$.

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 $\mathbf{\Gamma}$ by $\gamma\left(\theta\right) = \left(\gamma_i\left(\theta\right)\right)_{1\leq i\leq r}$ and to Λ by $\lambda\left(\theta\right) = \left(\lambda_i\left(\theta\right)\right)_{1\leq i\leq r}$. More generally, for a subset $A \subseteq \{1, 2, \ldots, r\}$, we define a projection of θ to a subset of $\mathbf{\Gamma}$ by $\gamma_A\left(\theta\right) = \left(\gamma_i\left(\theta\right)\right)_{i\in A}$. A particularly useful example of set A is

$$\{-i\} = \{1, \ldots, i-1, i+1, \cdots, r\},$$

for which $\gamma_{-i}\left(\theta\right) = \left(\gamma_1\left(\theta\right), \ldots, \gamma_{i-1}\left(\theta\right), \gamma_{i+1}\left(\theta\right), \gamma_r\left(\theta\right)\right)$. $\lambda_A\left(\theta\right)$ and $\lambda_{-i}\left(\theta\right)$ are defined similarly. We define the set $\Lambda_A = \{\lambda_A\left(\theta\right) : \theta \in \Theta\}$. 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 5. Define the mixture distribution

$$\bar{\mathbb{P}}_{(i,a,b,c)} = \frac{1}{D_{(i,a,b,c)}} \sum_{\theta \in \Theta_{(i,a,b,c)}} \mathbb{P}_\theta. \tag{52}$$

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*

$$\underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}}{\text{Average}} \left\{ \int \left(\frac{d\bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})}}{d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}} \right)^2 d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})} - 1 \right\} \le c_2^2, \tag{53}$$

then $\left\| \bar{\mathbb{P}}_{1,0} \wedge \bar{\mathbb{P}}_{1,1} \right\| \ge 1 - c_2 > 0$.

We can prove Lemma 15 using the same arguments as the proof of Lemma 8 in Cai and Zhou (2011). To complete the proof of Theorem 5 we need to verify only Equation (53).

5.2 Technical lemmas for proving Equation (53)

From the definition of $\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}$ in Equation (52) and $\theta = (\gamma,\lambda)$ with $\gamma = (\gamma_1,\cdots,\gamma_r)$ and $\lambda = (\lambda_1,\cdots,\lambda_r)$, $\gamma_1 = 0$ implies $\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}$ is a product of n multivariate normal distributions each with a covariance matrix,

$$\Sigma_{l,0} = \begin{pmatrix} 1 & \mathbf{0}_{1\times(p-1)} \\ \mathbf{0}_{(p-1)\times 1} & \mathbf{S}_{(p-1)\times(p-1)} \end{pmatrix} + (a_l - 1)\,\mathbf{I}_p, \text{ for } l = 1,2,\ldots,n, \tag{54}$$

where $\mathbf{S}_{(p-1)\times(p-1)} = (s_{ij})_{2\le i,j\le p}$ is uniquely determined by $(\gamma_{-1},\lambda_{-1}) = ((\gamma_2,...,\gamma_r),(\lambda_2,...,\lambda_r))$ 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 1's 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 (52) and $\theta = (\gamma,\lambda)$ with $\gamma = (\gamma_1,\cdots,\gamma_r)$ and $\lambda = (\lambda_1,\cdots,\lambda_r)$, $\bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})}$ is an average of $\binom{p_{\lambda_{-1}}}{k}$ number of products of multivariate normal distributions each with covariance matrix of the following form

$$\begin{pmatrix} 1 & \mathbf{r}_{1\times(p-1)} \\ \mathbf{r}_{(p-1)\times 1} & \mathbf{S}_{(p-1)\times(p-1)} \end{pmatrix} + (a_l - 1)\,\mathbf{I}_p, \text{ for } l = 1,2,\ldots,n, \tag{55}$$

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 (54). Note that the indices γ_i and λ_i are dropped from $\mathbf{r}$ and $\mathbf{S}$ to simplify the notations.

With Lemma 15 in place, it remains to establish Equation (53) in order to prove Lemma 14. 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*

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

Let $\Sigma_{l,i}$, $i = 1$ or 2, be two covariance matrices of the form (55). 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 (55). 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*

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

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

$$R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} = -\log\det\left(I - \Sigma_{l,0}^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})\right), \tag{56}$$

where $\Sigma_{l,0}$ is defined in (54) 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 notations. Define

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

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 (53) can be written as

$$\begin{aligned}
& \underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}}{\text{Average}} \left\{ \underset{\lambda_1,\lambda_1'\in\Lambda_1(\lambda_{-1})}{\text{Average}} \left[\exp\left(\frac{1}{2}\sum_{l=1}^{n} R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} \right) - 1 \right] \right\} \\
= & \underset{\lambda_1,\lambda_1'\in B}{\text{Average}} \left\{ \underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\lambda_1')}{\text{Average}} \left[\exp\left(\frac{1}{2}\sum_{l=1}^{n} R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} \right) - 1 \right] \right\},
\end{aligned} \tag{57}$$

where B is defined in Step 1.

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

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

Define

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

Lemma 18 *For $R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}$ defined in Equation (56) we have*

$$R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} = -2\log\left(1 - Ja_l^{-2}\epsilon_{n,p}^2\right) + \delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}, \tag{58}$$

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

$$\underset{(\lambda_1,\lambda_1')\in\Lambda_{1,J}}{\text{Average}} \left[\underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\lambda_1')}{\text{Average}} \exp\left(\frac{1}{2}\sum_{l=1}^{n} \delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} \right) \right] \le 3/2, \tag{59}$$

uniformly over all J defined in Lemma 17.

We will prove Lemma 18 in Section 5.4.

5.3 Proof of Equation (53)

We are now ready to establish Equation (53) using Lemma 18. It follows from Equation (58) in Lemma 18 that

$$\underset{\lambda_1,\lambda_1'\in B}{\text{Average}} \left\{ \underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\lambda_1')}{\text{Average}} \left[\exp\left(\tfrac{1}{2}\sum_{l=1}^{n} R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}} \right) - 1 \right] \right\} =$$

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

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

$$\mathbb{P}\left(\text{number of overlapping } \epsilon_{n,p}\text{'s} = J\right) = \binom{k}{J}\binom{p_{\lambda_{-1}}-k}{k-J}/\binom{p_{\lambda_{-1}}}{k} \leq \left(\frac{k^2}{p_{\lambda_{-1}}-k}\right)^J. \quad (60)$$

Equations (59) and (60) imply

$$\begin{aligned}
&\underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}}{\text{Average}}\left\{\int\left(\frac{d\bar{\mathbb{P}}_{(1,1,\gamma_{-1},\lambda_{-1})}}{d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}}\right)^2 d\bar{\mathbb{P}}_{(1,0,\gamma_{-1},\lambda_{-1})}-1\right\}\\
\leq\ & \sum_{J\geq 0}\frac{\binom{k}{J}\binom{p_{\lambda_{-1}}-k}{k-J}}{\binom{p_{\lambda_{-1}}}{k}}\left\{-\sum_{l=1}^{n}\log\left(1-J\epsilon_{n,p}^2/a_l^2\right)\right\}\frac{3}{2}-1\\
\leq\ & C\sum_{J\geq 1}\left(p^{\frac{\beta-1}{\beta}}\right)^{-J}\exp\left(2J\sum_{l=1}^{n}a_l^{-2}\cdot\frac{\upsilon^2\log p}{\sqrt{n}}\right)+1/2\\
\leq\ & C\sum_{J\geq 1}\left(p^{\frac{\beta-1}{\beta}}\right)^{-J}\exp\left(2Jc_\kappa\sqrt{n}\cdot\frac{\upsilon^2\log p}{\sqrt{n}}\right)+1/2\\
\leq\ & C\sum_{J\geq 1}\left(p^{\frac{\beta-1}{\beta}}\right)^{-J}\exp\left(2c_\kappa J\upsilon^2\log p\right)+1/2\leq C\sum_{J\geq 1}\left(p^{\frac{\beta-1}{2\beta}}\right)^{-J}+1/2<c_2^2,
\end{aligned}$$

where the third inequality is from (47), the fifth inequality is due to (46), and the last inequality is obtained by setting $c_2^2=3/4$. ∎

5.4 Proof of Lemma 18

Define

$$A_l=\left[I-a_l^{-2}\left(\Sigma_{l,0}-\Sigma_{l,1}\right)\left(\Sigma_{l,0}-\Sigma_{l,2}\right)\right]^{-1}\left(a_l^2\left(\Sigma_{l,0}\right)^{-2}-I\right)a_l^{-2}\left(\Sigma_{l,0}-\Sigma_{l,1}\right)\left(\Sigma_{l,0}-\Sigma_{l,2}\right),\quad (61)$$

and

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

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

$$\begin{aligned}
&R_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}}=\\
&-\log\det\left[I-a_l^{-2}\left(\Sigma_{l,0}-\Sigma_{l,1}\right)\left(\Sigma_{l,0}-\Sigma_{l,2}\right)-\left(a_l^2\Sigma_{l,0}^{-2}-I\right)a_l^{-2}\left(\Sigma_{l,0}-\Sigma_{l,1}\right)\left(\Sigma_{l,0}-\Sigma_{l,2}\right)\right]\\
&=-\log\det\left\{\left[I-A_l\right]\cdot\left[I-a_l^{-2}\left(\Sigma_{l,0}-\Sigma_{l,1}\right)\left(\Sigma_{l,0}-\Sigma_{l,2}\right)\right]\right\}\\
&=-\log\det\left[I-a_l^{-2}\left(\Sigma_{l,0}-\Sigma_{l,1}\right)\left(\Sigma_{l,0}-\Sigma_{l,2}\right)\right]-\log\det\left(I-A_l\right)\\
&=-2\log\left(1-J\epsilon_{n,p}^2/a_l^2\right)+\delta_{l,\lambda_1,\lambda_1'}^{\gamma_{-1},\lambda_{-1}},
\end{aligned}\quad (62)$$

where the last equation follows from Lemma 17.

Now we are ready to establish Equation (59). For simplicity we will write matrix norm $\|\cdot\|_2$ as $\|\cdot\|$ below. It is important to observe that $\mathrm{rank}\,(A_l) \le 2$ 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

$$\begin{aligned} &|\varrho_l| \le \|A_l\| \\ &\le \left\|a_l^2 \Sigma_{l,0}^{-2} - I\right\| \cdot a_l^{-2} \|\Sigma_{l,0} - \Sigma_{l,1}\| \|\Sigma_{l,0} - \Sigma_{l,2}\| / \left(1 - a_l^{-2} \|\Sigma_{l,0} - \Sigma_{l,1}\| \|\Sigma_{l,0} - \Sigma_{l,2}\|\right) \\ &\le \left(\left(\frac{3}{2}\right)^2 - 1\right) \frac{1}{3} \cdot \frac{1}{3} / \left(1 - \frac{1}{3} \cdot \frac{1}{3}\right) = 5/32 < 1/6, \end{aligned} \tag{63}$$

since $\left\|a_l^{-1}(\Sigma_{l,0} - \Sigma_{l,1})\right\| \le \left\|a_l^{-1}(\Sigma_{l,0} - \Sigma_{l,1})\right\|_1 = 2k\epsilon_{n,p} < 1/3$ and $\lambda_{\min}\left(a_l^{-1}\Sigma_{l,0}\right) \ge 1 - \left\|I - a_l^{-1}\Sigma_{l,0}\right\| \ge 1 - \left\|I - a_l^{-1}\Sigma_{l,0}\right\|_1 > 2/3$ from Equation (48).

Note that (63) and

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

imply

$$\delta_{l,\lambda_1,\acute{\lambda}_1}^{\gamma_{-1},\lambda_{-1}} \le 4\|A_l\|,$$

and then

$$\exp\left(\frac{1}{2}\sum_{l=1}^{n} \delta_{l,\lambda_1,\acute{\lambda}_1}^{\gamma_{-1},\lambda_{-1}}\right) \le \exp\left(2\sum_{l=1}^{n}\|A_l\|\right). \tag{64}$$

Since

$$\begin{cases} \left\|I - a_l^{-1}\Sigma_{l,0}\right\| \le \left\|I - a_l^{-1}\Sigma_{l,0}\right\|_1 = 2k\epsilon_{n,p} < 1/3 < 1, \\ \left\|a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})\right\| \le \frac{1}{3} \cdot \frac{1}{3} < 1, \end{cases} \tag{65}$$

we write

$$\begin{aligned} a_l^2\Sigma_{l,0}^{-2} - I &= \left(I - \left(I - a_l^{-1}\Sigma_{l,0}\right)\right)^{-2} - I = \left(I + \sum_{k=1}\left(I - a_l^{-1}\Sigma_{l,0}\right)^k\right)^2 - I \qquad (66) \\ &= \left[\sum_{m=0}^{\infty}(m+2)\left(I - a_l^{-1}\Sigma_{l,0}\right)^m\right]\left(I - a_l^{-1}\Sigma_{l,0}\right), \end{aligned}$$

where

$$\left\|\sum_{m=0}^{\infty}(m+2)\left(I - a_l^{-1}\Sigma_{l,0}\right)^m\right\| \le \sum_{m=0}^{\infty}(m+2)\left(\frac{1}{3}\right)^m < 3. \tag{67}$$

Define

$$A_{l*} = \left(I - a_l^{-1}\Sigma_{l,0}\right) \cdot a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2}). \tag{68}$$

From Equations (61) and (65)-(68) we have

$$\begin{aligned} \|A_l\| &\le \left\|\left[I - a_l^{-2}(\Sigma_{l,0} - \Sigma_{l,1})(\Sigma_{l,0} - \Sigma_{l,2})\right]^{-1}\right\| \left\|\sum_{m=0}^{\infty}(m+2)\left(I - a_l^{-1}\Sigma_{l,0}\right)^m\right\| \|A_{l*}\| \\ &< \frac{1}{1 - \frac{1}{3}\cdot\frac{1}{3}} \cdot 3 \cdot \|A_{l*}\| = \frac{27}{8}\|A_{l*}\| \le \frac{27}{8}\max\left\{\|A_{l*}\|_1, \|A_{l*}\|_\infty\right\}. \end{aligned}$$

The above result and (64) indicate that the proof of Lemma 18 is completed if we show

$$\underset{(\lambda_1,\lambda_1')\in\Lambda_{1,J}}{\text{Average}}\left[\underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\acute{\lambda_1})}{\text{Average}}\exp\left(\frac{27}{2}\sum_{l=1}^{n}\max\left\{\|A_{l*}\|_1,\|A_{l*}\|_\infty\right\}\right)\right]\le 3/2, \tag{69}$$

where $\|A_{l*}\|_1$ and $\|A_{l*}\|_\infty$ depend on the values of $\lambda_1, \acute{\lambda_1}$ and $(\gamma_{-1}, \lambda_{-1})$. We dropped the indices λ_1, $\acute{\lambda_1}$ and $(\gamma_{-1}, \lambda_{-1})$ from A_l to simplify the notations.

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 can not 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 1's in the upper triangular matrix by the construction of the parameter set, we thus have $n_{\lambda_{E_m}} \le r \cdot \frac{3}{4}$, which immediately implies $p_{\lambda_{E_m}} = r - n_{\lambda_{E_m}} \ge \frac{r}{4} \ge p/8 - 1$. Thus we have for every non-negative integer t

$$\begin{aligned}
&\mathbb{P}\left(\max\left\{\|A_{l*}\|_1,\|A_{l*}\|_\infty\right\}\ge 2t\cdot\epsilon_{n,p}\cdot k\epsilon_{n,p}^2\cdot a_l^{-3}\right)\\
\le\ &\mathbb{P}\left(\|A_{l*}\|_1\ge 2t\cdot\epsilon_{n,p}\cdot k\epsilon_{n,p}^2\cdot a_l^{-3}\right)+\mathbb{P}\left(\|A_{l*}\|_\infty\ge 2t\cdot\epsilon_{n,p}\cdot k\epsilon_{n,p}^2\cdot a_l^{-3}\right)\\
\le\ &2\sum_m \underset{\lambda_{E_m}}{\text{Average}}\frac{\binom{k}{t}\binom{p_{\lambda_{E_m}}}{k-t}}{\binom{p_{\lambda_{E_m}}}{k}}\le 2p\left(\frac{k^2}{p/8-1-k}\right)^t
\end{aligned}$$

from Equation (60), which immediately implies

$$\begin{aligned}
&\underset{(\lambda_1,\lambda_1')\in\Lambda_{1,J}}{\text{Average}}\left[\underset{(\gamma_{-1},\lambda_{-1})\in\Theta_{-1}(\lambda_1,\acute{\lambda_1})}{\text{Average}}\exp\left(\frac{27}{2}\sum_{l=1}^{n}\max\left\{\|A_{l*}\|_1,\|A_{l*}\|_\infty\right\}\right)\right]\\
\le\ &\exp\left(\frac{27}{2}\sum_{l=1}^{n}\frac{4\beta}{\beta-1}\cdot\epsilon_{n,p}\cdot k\epsilon_{n,p}^2\cdot a_l^{-3}\right)+\\
&\int_{\frac{2\beta}{\beta-1}}^{\infty}\left(27k\epsilon_{n,p}^3\sum_{l=1}^{n}a_l^{-3}\right)\exp\left(\frac{27}{2}\sum_{l=1}^{n}2t\cdot\epsilon_{n,p}\cdot k\epsilon_{n,p}^2\cdot a_l^{-3}\right)2p\left(\frac{k^2}{p/8-1-k}\right)^{t-1}dt\\
\le\ &\exp\left(54\cdot\left(\sum_{l=1}^{n}a_l^{-3}\right)\cdot\frac{\beta}{\beta-1}\cdot k\epsilon_{n,p}^3\right)\\
&+2p\int_{\frac{2\beta}{\beta-1}}^{\infty}\exp\left[(t+1)\cdot 27\left(\sum_{l=1}^{n}a_l^{-3}\right)k\epsilon_{n,p}^3-(t-1)\log\frac{p/8-1-k}{k^2}\right]dt.
\end{aligned} \tag{70}$$

Note that (47) implies

$$\sum_{l=1}^{n}a_l^{-3}\le\sum_{l=1}^{n}a_l^{-2}\le c_\kappa\sqrt{n},$$

using (11) and (44) we have

$$2\sqrt{n}k\epsilon_{n,p}^3\le\sqrt{n}\pi_n(p)\epsilon_{n,p}^{3-q}\le Mv^{3-q}n^{1/2}n^{(1-q)/4}(\log p)^{(q-3)/2}n^{(q-3)/4}(\log p)^{(3-q)/2}=Mv^{3-q},$$

and thus we can bound the first term on the right hand side of (70)

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

where the second inequality is from (45) and (46). We will show that the second term on the right hand side of (70) is negligible and hence establish (69). Indeed, since we have just shown that

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

the second term on the right hand side of (70) is bounded by

$$\begin{aligned}
& 2p\int_{\frac{2\beta}{\beta-1}}^{\infty} \exp\left[(t+1)\frac{\beta-1}{6\beta} - (t-1)\log\frac{p/8-1-k}{k^2}\right] dt \\
= \; & 2\left(\log\frac{p/8-1-k}{k^2} - \frac{\beta-1}{6\beta}\right)^{-1} \\
& \exp\left[\log p + \left(\frac{2\beta}{\beta-1}+1\right)\frac{\beta-1}{6\beta} - \left(\frac{2\beta}{\beta-1}-1\right)\log\frac{p/8-1-k}{k^2}\right] \\
= \; & O\left(p^{-1/\beta}[\log p]^{6/(\beta-1)+2}\right) = o(1),
\end{aligned}$$

where the second equality is from the fact that (11) and (44) together with $p \geq n^{\beta/2}$ indicate

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

and then

$$\begin{aligned}
\left(\frac{2\beta}{\beta-1}-1\right)\log\frac{p/8-k}{k^2} & \sim \left(\frac{2\beta}{\beta-1}-1\right)\log\left(pk^{-2}\right) \\
& \geq \left(\frac{2\beta}{\beta-1}-1\right)\left[\frac{\beta-1}{\beta}\log p + 3\log\log p - \log(Mv^{-2q}/4)\right] \\
& \sim \left(1+\frac{1}{\beta}\right)\log p. \; \blacksquare
\end{aligned}$$

Acknowledgements. Wang is partially supported by NSF grant DMS-105635, and Zhou is partially supported by NSF Career Award DMS-0645676 and NSF FRG grant DMS-0854975.

REFERENCES

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

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

Cai, T. and Zhou, H. (2011). Optimal rates of convergence for covariance matrix estimation. Under the review of Ann. Statist.

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

Callen, J., Govindaraj, S. and Xu, L. (2000). Large time and small noise asymptotic results for mean reverting processes with applications. *Economic Theory*, 16, 401-419.

El Karoui, N. (2008). Operator norm consistent estimation of large dimensional sparse covariance matrices. Ann. Statist. 36, 2717-2756.

Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of American Statistical Association* 102, 1349-1362.

Fan, J., Fan, Y. and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. Journal of Econometrics 147, 186-197.

Fan, J., Li, Y., and Yu, K. (2011). Vast volatility matrix estimation using high frequency data for portfolio selection. Technical report.

Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I. local asymptotic normality. ESAIM: Probability and Statistics 5, 225-242.

Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. II. optimal estimators. ESAIM 5, 225-242.

He, S. W., Wang, J. G., and Yan, J. A. (1992). Semimartingale Theory and Stochastic Calculus. Science Press and CRC Press Inc., Beijing.

Hörmander, L. (1983). The Analysis of Linear Partial Differential Operators I, Grundl. Math. Wissenschaft., 256, Springer.

Huang, J., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. Biometrika 93, 85-98.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. Ann. Statist. **29**, 295-327.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal component analysis in high dimensions (with discussions). J. Amer. Statist. Assoc. 104, 682-703.

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

Le Cam, L. (1986). Asymptotic Methods in Statistical Decision Theory. Springer-Verlag, New York.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empirical Finance **10**, 603-621.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. J. Multi. Anal. **88**, 365-411.

Mueschke, N. J. and Andrews, M. J. (2006). Investigation of scalar measurement error in diffusion and mixing processes. *Experiments in Fluids*, 40, 165-175.

Meyer, C. D. (2004). Matrix Analysis and Applied Linear Algebra. SIAM

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

Saulis, L. and Statulevičius, V. (1991). Limit Theorems for Large Deviations. Kluwer Academic Publishers, Dordrecht.

Salkuyeh, D. K.(2006). Positive integer powers of the tridiagonal Toeplitz matrices. International Mathematical Forum 1, 1061 - 1065

Tao, M., Wang, Y., Yao, Q. and Zou, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. Journal of the American Statistical Association 106, 1025-1040.

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

Whitmore, G. A. (1995). Estimating degradation by a Wiener diffusion process subject to measurement error. *Lifetime Data Analysis*, 1, 307-319.

Wilkinson, J.H. (1988). The Algebraic Eigenvalue Problem. Oxford University Press, USA.

Wu, W. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. Biometrika **90**, 831-844.

Yu, B. (1997). Assouad, Fano, and Le Cam. Festschrift for Lucien Le Cam. D. Pollard, E. Torgersen, and G. Yang (eds), 423-435, Springer-Verlag, New York.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. Biometrika, 94, 19-35.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394-1411.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12, 1019-1043.

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

Zheng, X. and Li, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. Technical report.

Minjing Tao and Yazhen Wang
Department of Statistics
University of Wisconsin-Madison
Madison, WI 53706

Harrison H. Zhou
Department of Statistics
Yale University
New Haven, CT 06511