Submitted to the Annals of Statistics

ON THE SYSTEMATIC AND IDIOSYNCRATIC VOLATILITY WITH LARGE PANEL HIGH-FREQUENCY DATA

By Xin-Bing KONG

Nanjing Audit University

In this paper, we separate the integrated (spot) volatility of an individual Itô process into integrated (spot) systematic and idiosyncratic volatilities, and estimate them by aggregation of local factor analysis (localization) with large-dimensional high-frequency data. We show that, when both the sampling frequency n and the dimensionality p go to infinity and $p \geq C\sqrt{n}$ for some constant C, our estimators of the integrated (spot) systematic and idiosyncratic volatilities are $\sqrt{n}$ ($n^{1/4}$ for spot estimates) consistent, the best rate achieved in estimating the integrated (spot) volatility which is readily identified even with univariate high-frequency data. However, when $Cn^{1/4} \leq p < C\sqrt{n}$, aggregation of $n^{1/4}$-consistent local estimates of systematic and idiosyncratic volatilities results in p-consistent (not $\sqrt{n}$-consistent) estimates of integrated systematic and idiosyncratic volatilities. Even more interestingly, when $p < Cn^{1/4}$, integrated estimate has the same convergence rate as the spot estimate, both being p-consistent. This reveals a distinctive feature from aggregating local estimates in the low-dimensional high-frequency data setting. We also present estimators of the integrated (spot) idiosyncratic volatility matrices as well as their inverse matrices under some sparsity assumption. We finally present a factor-based estimator of the inverse of the spot volatility matrix. Numerical studies including the Monte-Carlo experiments and real data analysis justify the performance of our estimators.

1. Introduction. Itô processes are widely used in finance to model the price dynamics of assets, cf, Barndorff-Nielsen (2002), Mykland and Zhang (2009) and the references therein. Since the integrated (spot) volatility of a Itô process is an important characteristics in risk analysis, asset allocation, and derivatives pricing (e.g., variance swap), statistical inference on it becomes a hot topic recently, cf, Andersen *et al.* (2003), Kong *et al.* (2015), Chen and Xu (2014) and Jacod and Todorov (2014). Extensions to integrated volatility matrix of fixed dimension in-

*Kong's work is partly supported by NSFC 11201080, 11571250, HSSYF of Chinese Ministry of Education (12YJC910003), and PAPD of Jiangsu Higher Education Institution.

AMS 2000 subject classifications: Primary 62M05, 62G20;secondary60J75, 60G20

Keywords and phrases: High dimensional Itô process; Common driving process; Specific driving process

cludes Barndorff-Nielsen *et al.* (2008) and Christensen *et al.* (2010). Inspired by the risk decomposition into a systematic part and an idiosyncratic part, in this paper, we separate the volatility (integrated and spot) into systematic and specific (idiosyncratic) volatilities. This would have potential applications in finance. For example, estimates of the integrated systematic and idiosyncratic volatilities would help to get a pricing decomposition for the variance swap. The estimates are also important in risk management and portfolio analysis. It is well known in factor analysis that the communality and the specific variance can not be identified using univariate observations or multivariate observations of fixed dimension. Therefore in this paper, we will estimate those two by using large panel high-frequency data, that is when both the dimensionality and the sample size are large enough, which is a common feature in modern data base.

Partly due to the development of data collection technology, we are now facing vast portfolios. Large-dimensional Itô process becomes a natural tool to model the prices of a large pool of assets with serial and cross-sectional dependence. Consider the following large-dimensional Itô process defined on a filtered probability space $\{\Omega, \mathcal{F}, \mathcal{F}_t, P\}$,

$$X_{it} = X_{i0} + \int_0^t \mu_{is} ds + \int_0^t \sigma_{is}^1 dW_s^1 + ... + \int_0^t \sigma_{is}^r dW_s^r + \int_0^t \sigma_{is}^* dW_{is}^*, \ 1 \le i \le p, \tag{1.1}$$

where X_{it} stands for the log price trajectory of asset i, X_{i0} is the initial value, μ_i's, σ_i^l's $(1 \le l \le r)$, σ_i^*'s are locally bounded adapted processes, and $\boldsymbol{W} = (W^1, ..., W^r)'$ is an r-dimensional Brownian motion and $\boldsymbol{W}^* = (W_1^*, ..., W_p^*)'$ is a p-dimensional Brownian motion with correlation matrix $\boldsymbol{\rho}^*$. In matrix form, (1.1) can be rewritten as

$$d\boldsymbol{X}_t = \boldsymbol{B}_t dt + \boldsymbol{\sigma}_t d\boldsymbol{W}_t + \boldsymbol{\sigma}_t^* d\boldsymbol{W}_t^*,$$

where $\boldsymbol{B}_t = (\mu_{1t}, ..., \mu_{pt})'$, $\boldsymbol{X}_t = (X_{1t}, ..., X_{pt})'$, $\boldsymbol{\sigma}_t^* = diag(\sigma_{1t}^*, ..., \sigma_{pt}^*)$, and

$$\boldsymbol{\sigma}_t = \begin{pmatrix} \sigma_{1t}^1 & \cdots & \sigma_{1t}^r \\ \sigma_{2t}^1 & \cdots & \sigma_{2t}^r \\ & \vdots & \\ \sigma_{pt}^1 & \cdots & \sigma_{pt}^r \end{pmatrix}.$$

For reason of identifiability, we assume throughout the paper that $corr(\boldsymbol{W}) = I_r$. We also assume that $\boldsymbol{W}$ and $\boldsymbol{W}^*$ are independent although it can be relaxed to allow for weak dependence.

Model (1.1) assumes that the log price of each asset is driven by r common Brownian motions and an asset-specific Brownian motion. In the simulation studies, Barndorff-Nielsen *et al.* (2008) and Fan *et al.* (2012) used similar versions of (1.1) to mimic the real log prices. Interestingly, Aït-Sahalia and Xiu (2016b) and Pelger (2016) recently specify $\boldsymbol{X}_t$ as

$$\boldsymbol{X}_t = \boldsymbol{\Lambda} \boldsymbol{Y}_t + \boldsymbol{Z}_t, \tag{1.2}$$

where $\boldsymbol{\Lambda}$ is a $p \times r$ constant factor loading matrix, $\boldsymbol{Y}$ is an r-dimensional unobservable common factor process, and $\boldsymbol{Z}$ is a p-dimensional process representing the idiosyncratic component. Both $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are multivariate Itô processes. The stringent condition of model (1.2) is the constancy of $\boldsymbol{\Lambda}$. For example, if $r = 1$, then the component volatility processes of $\boldsymbol{\Lambda}\boldsymbol{Y}_t$ are perfectly correlated which is not realistic. In contrast, model (1.1) even allows for independence among component volatility processes of the systematic part.

Under model (1.1), the spot systematic and idiosyncratic volatility matrices are, respectively,

$$\boldsymbol{\Sigma}_s^c = \boldsymbol{\sigma}_s \boldsymbol{\sigma}_s', \quad \text{and} \quad \boldsymbol{\Sigma}_s^e = \boldsymbol{\sigma}_s^* \boldsymbol{\rho}^* \boldsymbol{\sigma}_s^*.$$

The integrated systematic and idiosyncratic volatilities for asset i are, respectively,

$$\Gamma_{ii}^c = \int_0^t \boldsymbol{\sigma}_{is}' \boldsymbol{\sigma}_{is} ds = \int_0^t \Sigma_{ii}^c(s) ds, \text{ and } \Gamma_{ii}^e = \int_0^t \sigma_{is}^{*2} ds = \int_0^t \Sigma_{ii}^e(s) ds, \tag{1.3}$$

where $\boldsymbol{\sigma}_{is}' = (\sigma_{is}^1, ..., \sigma_{is}^r)$, and $\Sigma_{ii}^c(s)$ and $\Sigma_{ii}^e(s)$ are the ith diagonal entries of $\boldsymbol{\Sigma}_s^c$ and $\boldsymbol{\Sigma}_s^e$, respectively. The sum of these two is exactly the integrated volatility for X_i.

There are many discrete analogies of model (1.1) popularly adopted in econometrics. Ross (1976) introduced the arbitrage pricing theory (APT) model with strict factor structure. Chamberlain and Rothschild (1983) extended the APT model to an approximate factor model allowing for non-zero off-diagonal elements in the residual covariance matrix. Motivated by them, we also impose a sparse structure on $\boldsymbol{\rho}^*$, which naturally renders a sparse structure of the integrated idiosyncratic volatility matrix

$$\boldsymbol{\Gamma}^e = (\Gamma_{im}^e)_{p\times p} = \left(\int_0^T \sigma_{is}^* \rho_{im}^* \sigma_{ms}^* ds \right)_{p\times p}, \ \textit{with}\ \rho_{im}^* = \boldsymbol{\rho}^*(i, m), \tag{1.4}$$

see Assumption 1 below.

The first estimator of large sparse covariation matrix using high-frequency data dates back to Wang and Zou (2010) by using universal thresholding technique. Their estimator was later refined and extensively investigated by Tao *et al.* (2013a) and (2013b). Kim *et al.* (2015) introduced an adaptive thresholding version of Wang and Zou (2010)'s estimator. The thresholding and adaptive thresholding techniques dealing with i.i.d data were invented, respectively, by Bickel and Levina (2008) and Cai and Liu (2011). However, all these works imposed a sparsity assumption on the integrated volatility matrix, which is rarely met for model (1.1). But rather their sparsity assumptions are especially suited and well interpretable for the idiosyncratic volatility matrix, hence in this paper we assume instead that $\boldsymbol{\rho}^*$ is sparse.

Related to the portfolio selection theory, it is of vital importance to estimate the inverse of the spot volatility matrix, that is, the inverse of

$$\boldsymbol{\Sigma}_s \equiv \boldsymbol{\Sigma}_s^c + \boldsymbol{\Sigma}_s^e. \tag{1.5}$$

In Fan *et al.* (2012), in principal, $\tau\mathbf{\Sigma}_s$ was used as a proxy of the covariance matrix $\int_s^{s+\tau} E_{\mathcal{F}_s}\mathbf{\Sigma}_u du$ in a short future holding period $(s, s+\tau)$, and hence the optimal allocation vector relies on the inverse of $\mathbf{\Sigma}_s$.

In this paper, we provide estimators for the integrated (spot) systematic and idiosyncratic volatilities for an individual Itô process. The convergence rate of our estimators achieves $\sqrt{n}$ ($n^{1/4}$ for spot estimates) as in estimating the integrated (spot) volatility if $p \geq C\sqrt{n}$ ($p \geq Cn^{1/4}$ for spot estimates) for some constant C. We later present estimators of the integrated and spot idiosyncratic volatility matrices as well as their inverse matrices. The estimators converge in the operator norm at the rate of $s_0(p)(\frac{1}{\sqrt{p}} + \frac{\sqrt{\log p}}{\sqrt{n}})^{1-q}$ (or $s_0(p)(\frac{1}{\sqrt{p}} + \frac{\sqrt{\log p}}{n^{1/4}})^{1-q}$ for spot estimates) where $s_0(p)$ is a measure of sparsity and q is some constant in $[0, 1)$.

To deal with the time varying feature of the volatility processes, our basic technique is based on local factor analysis and then aggregate the local communalities and specific variances to form integrated versions. Aggregation of local analysis of high-frequency data is natural and indeed not new, and it arises frequently in estimation of volatility functionals, cf, Mykland and Zhang (2009), Jacod and Rosenbaum (2013), Jacod and Todorov (2014), Mykland *et al.* (2012), and Todorov and Tauchen (2012a), (2012b). However, the aggregation of factor analysis for large-dimensional Itô process is still unclear in the high-frequency data literature. Indeed, the asymptotic results demonstrate that while the spot estimators partly depend on $\frac{1}{n^{1/4}}$, the integrated estimators partly depend on a higher rate of $\frac{1}{\sqrt{n}}$, and that the convergence rate of all estimates relies on $\frac{1}{\sqrt{p}}$ due to the cross-section dependence of the large-dimensional idiosyncratic driving Brownian motion $\boldsymbol{W}^*$. The most interesting feature of aggregation is the joint dependence on p and n in estimating the individual systematic and idiosyncratic volatilities. See Theorem 1 and Remark 1 below for details. This shows a distinct scenario from the typical low-dimensional setting, where aggregation of $n^{1/4}$-consistent local estimates results in $\sqrt{n}$ consistent global estimates.

We finally present a factor-based estimator of the inverse of the spot volatility matrix which converges to the population counterpart with the rate of $s_0(p)(\frac{1}{\sqrt{p}} + \frac{\sqrt{\log p}}{n^{1/4}})^{1-q} + \frac{1}{\sqrt{p}} + 1/n^{1/4} + \frac{1}{n^{1/8}p^{3/8}}$, where the term $1/n^{1/4}$ is due to the discretization error within a local window. The implication is that the discretization error does matter, but it is under control in practice if the period length is short and thus the estimated inverse of the spot matrix is still reliable. We failed in proving the consistency of the inverse of the factor-based thresholding estimator of the integrated volatility matrix under the general model (1.1) and typical assumptions below. The major difficulty is that the sum of all estimates of spot systematic volatility matrices has no spiked eigenvalues, see Remark 4 for more discussions.

Related to this work, there are some interesting reference papers. Aït-Sahalia and Xiu (2016a) constructed estimators of realized eigenvalues, eigenvectors, and principal components using high-frequency data. Our paper also needs the realized instantaneous eigenvectors as an interim step to estimate the systematic volatilities and factor scores. The difference is that, in their paper, they are interested in spectral functionals of multivariate Itô process of fixed dimension, but in our paper we need $p \to \infty$ and $n \to \infty$ simultaneously. Aït-Sahalia and Xiu (2016b) proposed factor-based estimators of the integrated volatility matrices and proved the asymptotic properties. But their theory is carried out under model (1.2) and the constancy assumption ensures the rationale of global PCA in their paper. Contrast to their paper, we developed the theory under the more general model (1.1) and we used the local PCA technique adapted to the time varying volatilities. Besides, we provided the asymptotics of the estimates for the individual integrated systematic and idiosyncratic volatilities, as well as that for the spot systematic and idiosyncratic volatilities. Fan *et al.* (2016) constructed a factor-based covariation matrix estimator of large-dimensional Itô process driven by Brownian motions using high-frequency data, but they assumed that the time-continuous factor processes as well as the number of factor processes are observable as in Fama-French's factor model. In the present paper, we assume latent common factors and unknown number of common driving Brownian motions. Bai and Ng (2003), and Fan *et al.* (2013) introduced novel approaches to do statistical inferences on discrete approximate factor model in contrast to the continuous-time counterpart (1.1). Moreover, our inference is based on high-frequency data due to absence of long term stationarity of X_i's and the time varying feature of volatility processes, instead of the discrete stationary panel data. Or mathematically, our asymptotic regime is of the infill type with fixed time horizon while theirs let the time horizon go to infinity and the time lag fixed. Interesting papers on large covariance matrix for discrete time models also include Li and Chen (2012).

The rest of the paper is arranged as follows. In section 2, we give a heuristic introduction to our approach. In section 3, we give some technical assumptions. We present our main results in section 4. Section 5 is devoted to numerical studies, including monte carlo simulations and real data analysis. Section 6 concludes. All technical proofs are relegated in the appendix or supplementary materials.

2. Methodology. In this section, we give a heuristic introduction of the methodology. We first assume that r is known. We also assume that the data set is discretely sampled from the large-dimensional Itô process, $\boldsymbol{X} = (X_1, ..., X_p)'$, with equal sampling frequency, $\Delta_n = T/n$, where T is a fixed time horizon and n is the sample size. Mathematically, we will consider the asymptotic regime that $\Delta_n \to 0$. This sample scheme is also adopted in Fan *et al.* (2016), Jacod and Podolskij (2013) and many references therein assuming absence of microstructure noise. The data set sampled sparsely in minutes in the real data section belongs to this class. We avoid using ultra-high frequency data simply for reducing the bias caused by the microstructure noise. As demonstrated empirically in Aït-Sahalia *et al.* (2005) and Fan *et al.* (2016), sampling in minutes leads to satisfactory results. For dense high frequency data with additive noises, standard denoising techniques (e.g., the pre-averaging method in Jacod *et al.* (2009)) are efficient. A hybrid of the aggregation of local factor analysis and the pre-averaging technique is expected to work for the dense high-frequency data scenario, but it requires more complicated technicalities and assumptions to guarantee the feasibility of local factor analysis (PCA). So, for simplicity, we choose to start from the simple sparse sampling setting assuming noise free as in Jacod and Podolskij (2013), Fan *et al* (2016) and Aït-Sahalia and Xiu (2016a) (2016b). More discussions can be found in Section 6.

Let X_{it_k} be the observation of X_i at time t_k with $\Delta_n = t_k - t_{k-1}$, and $\Delta_j^n X_i = X_{it_j} - X_{it_{j-1}}$. Let $\boldsymbol{\delta}_s = (\Delta^n_{\lceil \frac{s}{\Delta_n}\rceil + j} X_i)_{i=1,...,p}^{j=1,...,k_n} / \sqrt{\Delta_n} \equiv (\delta^s_{ij})_{p\times k_n}$ be a $p \times k_n$ matrix, where $\lceil x \rceil$ stands for the smallest integer no smaller than x. $\boldsymbol{\mu}_s = (\mu_{it_{\lceil \frac{s}{\Delta_n}\rceil + j}})_{i=1,...,p}^{j=1,...,k_n}$ a $p \times k_n$ matrix, $\boldsymbol{F}_s = (\Delta^n_{\lceil \frac{s}{\Delta_n}\rceil + j} W^l)_{l=1,...,r}^{j=1,...,k_n} / \sqrt{\Delta_n}$ an $r \times k_n$ matrix, $\boldsymbol{\sigma}_s = (\sigma^l_{is})_{i=1,...,p}^{l=1,...,r}$ a $p \times r$ matrix, $\boldsymbol{\sigma}^*_s = diag\{\sigma^*_{1s}, ..., \sigma^*_{ps}\}$ a $p \times p$ matrix, and

$$\boldsymbol{F}^*_s = (\Delta^n_{\lceil \frac{s}{\Delta_n}\rceil + j} W^*_i)_{i=1,...,p}^{j=1,...,k_n} / \sqrt{\Delta_n} \equiv (\boldsymbol{F}^*_s(1), ..., \boldsymbol{F}^*_s(k_n))$$

a $p \times k_n$ matrix. Then as $k_n\Delta_n$ shrinks to zero, we expect that

$$\boldsymbol{\delta}_s \approx \boldsymbol{\mu}_s \sqrt{\Delta_n} + \boldsymbol{\sigma}_s \boldsymbol{F}_s + \boldsymbol{\sigma}^*_s \boldsymbol{F}^*_s \equiv \boldsymbol{\mu}_s \sqrt{\Delta_n} + \bar{\boldsymbol{\delta}}_s. \tag{2.1}$$

Now the right hand side of (2.1) has exactly the discrete approximate factor structure with $\boldsymbol{\mu}_s\sqrt{\Delta_n}$ a negligible mean, $\boldsymbol{\sigma}_s$ the factor loading matrix "fixed" in a block of length $k_n\Delta_n$ but varies across blocks, $\boldsymbol{F}_s$ the common factors, $\boldsymbol{F}^*_s$ and $\boldsymbol{\sigma}^*_s$ the specific factors and their loadings.

Now within the local window $(s, \lceil \frac{s}{\Delta_n} \rceil \Delta_n + k_n\Delta_n)$, we do principal component analysis for $\frac{\boldsymbol{\delta}'_s \boldsymbol{\delta}_s}{pk_n}$. Let $\boldsymbol{V}_s = diag\{v^1_s, ..., v^r_s\}$ (v^l_s's decreasing) and $\hat{\boldsymbol{F}}_s = (\hat{\boldsymbol{F}}_s(1), ..., \hat{\boldsymbol{F}}_s(k_n))$ be, respectively, the diagonal matrix consisting of the eigenvalues and $\sqrt{k_n}$ times the eigenvector matrix of $\frac{\boldsymbol{\delta}'_s \boldsymbol{\delta}_s}{pk_n}$, with v^l_s corresponding to the eigenvector $\hat{\boldsymbol{F}}'_{sl}/\sqrt{k_n}$, the lth row of $\hat{\boldsymbol{F}}_s/\sqrt{k_n}$. Ignoring the discritization error and the drift term $\boldsymbol{\mu}_s\sqrt{\Delta_n}$, by Bai (2003) or Fan *et al.* (2013), we have

$$\hat{\boldsymbol{F}}_s(j) \approx \boldsymbol{H}_s \boldsymbol{F}_s(j), \quad \hat{\boldsymbol{\sigma}}_s \equiv \frac{\boldsymbol{\delta}_s \hat{\boldsymbol{F}}'_s}{k_n} \approx \boldsymbol{\sigma}_s \boldsymbol{H}_s^{-1}, \tag{2.2}$$

where $\boldsymbol{H}_s = \frac{\boldsymbol{V}_s^{-1} \hat{\boldsymbol{F}}_s \boldsymbol{F}'_s \boldsymbol{\sigma}'_s \boldsymbol{\sigma}_s}{pk_n}$. From (2.2), though $\boldsymbol{F}_s$ and $\boldsymbol{\sigma}_s = (\boldsymbol{\sigma}_{1s}, ..., \boldsymbol{\sigma}_{ps})'$ are not identified themselves, their product is indeed identifiable and $\hat{\boldsymbol{\sigma}}_s \hat{\boldsymbol{F}}_s \approx \boldsymbol{\sigma}_s \boldsymbol{F}_s$. This motivates us to estimate the spot systematic and idiosyncratic volatilities, respectively, by

$$\begin{aligned} \hat{\Sigma}^c_{im}(s,r) &= \hat{\boldsymbol{\sigma}}'_{is} \hat{\boldsymbol{\sigma}}_{ms}, \quad \hat{\Sigma}^e_{ii}(s,r) = \frac{1}{k_n}\sum_{j=1}^{k_n} (\delta^s_{ij})^2 - \hat{\Sigma}^c_{ii}(s,r) \\ \hat{\Sigma}^e_{im}(s,r) &= \frac{1}{k_n}\sum_{j=1}^{k_n} (\delta^s_{ij} - \hat{\boldsymbol{\sigma}}'_{is}\hat{\boldsymbol{F}}_s(j))(\delta^s_{mj} - \hat{\boldsymbol{\sigma}}'_{ms}\hat{\boldsymbol{F}}_s(j)), \ i \neq m. \end{aligned} \tag{2.3}$$

Before presenting the estimators of the integrated systematic and idiosyncratic volatilities, we introduce some more notations. We separate the sampling time points $\{0 = t_0 < t_1 < ... < t_n = T\}$ into $[\frac{n}{k_n}]$ ($[x]$ stands for the largest integer no larger than x) non-overlapping blocks with each block containing k_n increments. Throughout the paper, we assume that $k_n/\sqrt{n}$ is bounded from above and from below. For simplicity, we denote $\boldsymbol{\mu}_{(k-1)k_n\Delta_n}$, $\boldsymbol{\sigma}_{(k-1)k_n\Delta_n}$, $\boldsymbol{\sigma}^*_{(k-1)k_n\Delta_n}$, $\boldsymbol{F}_{(k-1)k_n\Delta_n}$, $\hat{\boldsymbol{F}}_{(k-1)k_n\Delta_n}$, $\boldsymbol{F}^*_{(k-1)k_n\Delta_n}$, $v^l_{(k-1)k_n\Delta_n}$, $\hat{\boldsymbol{\sigma}}_{t_{(k-1)k_n}}$, $\boldsymbol{H}_{t_{(k-1)k_n}}$ and $\boldsymbol{\delta}_{(k-1)k_n\Delta_n}$ by $\boldsymbol{\mu}_k$, $\boldsymbol{\sigma}_k$, $\boldsymbol{\sigma}^*_k$, $\boldsymbol{F}_k$, $\hat{\boldsymbol{F}}_k$, $\boldsymbol{F}^*_k$, v^k_l, $\hat{\boldsymbol{\sigma}}_k$, $\boldsymbol{H}_k$ and $\boldsymbol{\delta}_k$, respectively.

Then a natural way to estimate the integrated systematic and idiosyncratic (co)-volatilities is to aggregate the local estimates as

$$\begin{aligned} \hat{\Gamma}^c_{im}(r) &= \sum_{k=1}^{[n/k_n]} k_n\Delta_n \hat{\Sigma}^c_{im}(t_{(k-1)k_n}, r), \\ \hat{\Gamma}^e_{im}(r) &= \sum_{k=1}^{[n/k_n]} k_n\Delta_n \hat{\Sigma}^e_{im}(t_{(k-1)k_n}, r), \end{aligned} \tag{2.4}$$

where the notation $A(r)$ emphasizes the dependence on r, and $\hat{\Sigma}^c_{im}(t_{(k-1)k_n}, r)$ and $\hat{\Sigma}^e_{im}(t_{(k-1)k_n}, r)$ are, respectively, the im-th entry of $\hat{\boldsymbol{\Sigma}}^c_{t_{(k-1)k_n}}(r)$ and $\hat{\boldsymbol{\Sigma}}^e_{t_{(k-1)k_n}}(r)$.

Next, we present the estimators of the integrated and spot idiosyncratic volatility matrices under the following sparse condition.

Assumption 1. *$\boldsymbol{\rho}^*$ belongs to*

$$\mathcal{U}_q(s_0(p)) = \{\boldsymbol{\rho}^*; \max_m \sum_{i=1}^p |\rho^*_{im}|^q \leq s_0(p)\},$$

where $0 \leq q < 1$, and $s_0(p)$ is some function of p.

For the m-dependent situation, $s_0(p)$ defined in Assumption 1 is bounded, but for the general weakly dependent structure, $s_0(p)$ may go to infinity slowly. When $q = 0$, Assumption 1 means that each asset-specific factor is correlated to at most $s_0(p)$ other asset specific factors. This is realistic in financial theory and empirical studies, cf, Fan *et al.* (2016).

First, we estimate the spot covariations by (2.3). Then our thresholding estimator of the spot idiosyncratic volatility matrix is

$$\hat{\boldsymbol{\Sigma}}^{e\mathcal{T}}_s(r) \equiv \left(\hat{\Sigma}^{e\mathcal{T}}_{im}(s,r)\right)_{p\times p}, \ \textit{with} \ \hat{\Sigma}^{e\mathcal{T}}_{im}(s,r) = \begin{cases} g_\lambda(\hat{\Sigma}^e_{im}(s,r)) & i \neq m \\ \hat{\Sigma}^e_{ii}(s,r) & i = m, \end{cases}$$

where $g_\lambda(z)$ is a class of thresholding functions satisfying (i) $|g_\lambda(z)| \leq c|y|$ for all z, y with $|z - y| \leq \lambda$, and some $c > 0$; (ii) $g_\lambda(z) = 0$ for $|z| \leq \lambda$; (iii) $|g_\lambda(z) - z| \leq \lambda$, for all $z \in R$.

These three conditions are satisfied by, for example, the hard thresholding function $g_\lambda(z) = zI(|z| > \lambda)$, the soft thresholding rule $g_\lambda(z) = sgn(z)(z-\lambda)_+$ and the adaptive lasso rule $g_\lambda(z) = z(1 - |\lambda/z|^\eta)_+$ with $\eta \geq 1$.

Analogously, our thresholding estimator of the integrated idiosyncratic volatility is

$$\hat{\mathbf{\Gamma}}^{e\mathcal{T}}(r) \equiv \left(\hat{\Gamma}^{e\mathcal{T}}_{im}(r)\right)_{p\times p}, \ with \ \hat{\Gamma}^{e\mathcal{T}}_{im}(r) = \begin{cases} g_\lambda(\hat{\Gamma}^e_{im}(r)) & i \neq m \\ \hat{\Gamma}^e_{ii}(r) & i = m. \end{cases}$$

Finally, we consider estimating the inverse of $\mathbf{\Sigma}_s$. Our factor-based estimator is $\hat{\mathbf{\Sigma}}_s^{-1}(r)$, the inverse of $\hat{\mathbf{\Sigma}}_s(r)$, where

$$\hat{\mathbf{\Sigma}}_s(r) = \hat{\mathbf{\Sigma}}^c_s(r) + \hat{\mathbf{\Sigma}}^{e\mathcal{T}}_s(r) \equiv (\hat{\Sigma}^c_{im}(s,r))_{p\times p} + \hat{\mathbf{\Sigma}}^{e\mathcal{T}}_s(r).$$

Now we consider the situation when r is unknown. In this case, we simply replace r by an estimated number $\hat{r}$. In this paper, we use the approach studied in Kong (2016a) by minimizing the penalized aggregated mean squared residual error, that is, for some bounded r_m,

$$\hat{r} = arg \min_{l\leq r_m} U(l, \tilde{\boldsymbol{F}}^l) + \beta l g(p,n) \equiv arg \min_{l \leq r_m} \frac{1}{[n/k_n]} \sum_{k=1}^{[n/k_n]} V(l, \tilde{\boldsymbol{F}}^l_k) + \beta l g(p,n),$$

where $\tilde{\boldsymbol{F}}^l = (\tilde{\boldsymbol{F}}^l_1, ..., \tilde{\boldsymbol{F}}^l_{[n/k_n]})$ with $\tilde{\boldsymbol{F}}^l_k = \check{\boldsymbol{F}}^l_k \boldsymbol{\delta}'_k \boldsymbol{\delta}_k$,

$$V(l, \tilde{\boldsymbol{F}}^l_k) = \frac{1}{pk_n} tr\left(\boldsymbol{\delta}_k (I_{k_n} - \tilde{\boldsymbol{F}}^{l\prime}_k (\tilde{\boldsymbol{F}}^l_k \tilde{\boldsymbol{F}}^{l\prime}_k)^{-1} \tilde{\boldsymbol{F}}^l_k) \boldsymbol{\delta}'_k\right),$$

β is a tuning parameter, and $g(p,n)$ is a function depending on p and n satisfying $(\sqrt{p} \wedge \sqrt{k_n})^2 g(p,n) \to \infty$ and $g(p,n) \to 0$ as $p, n \to \infty$. Here $\check{\boldsymbol{F}}^l_k$ is a $l \times k_n$ matrix whose jth row is $\sqrt{k_n}$ times the eigenvector of $\boldsymbol{\delta}'_k \boldsymbol{\delta}_k$ corresponding to v^k_j for $1 \leq j \leq l$. Theorem 3 in Kong (2015a) shows that

$$\lim_{p,n\to\infty} P(\hat{r} = r) = 1. \tag{2.5}$$

So our final estimators are $\hat{\Gamma}^c_{ii}(\hat{r})$, $\hat{\Sigma}^c_{ii}(s,\hat{r})$, $\hat{\Gamma}^e_{ii}(\hat{r})$, $\hat{\Sigma}^e_{ii}(s,\hat{r})$, $\hat{\mathbf{\Gamma}}^{e\mathcal{T}}(\hat{r})$, $\hat{\mathbf{\Sigma}}^{e\mathcal{T}}_s(\hat{r})$, and $\hat{\mathbf{\Sigma}}^{-1}_s(\hat{r})$, the plug-in versions of $\hat{\Gamma}^c_{ii}(r)$, $\hat{\Sigma}^c_{ii}(s,r)$, $\hat{\Gamma}^e_{ii}(r)$, $\hat{\Sigma}^e_{ii}(s,r)$, $\hat{\mathbf{\Gamma}}^{e\mathcal{T}}(r)$, $\hat{\mathbf{\Sigma}}^{e\mathcal{T}}_s(r)$, and $\hat{\mathbf{\Sigma}}^{-1}_s(r)$, respectively.

3. Setup Assumptions. To establish the theoretical results, we need the following technical assumptions. The first assumption gives some regularity conditions on the coefficient processes of the Itô processes. They are commonly used in the literature, cf, Jacod and Todorov (2014), Jing *et al.* (2012a), Jing *et al.* (2012b), Kong *et al.* (2015a) for univariate models, and Wang and Zou (2010), Fan *et al.* (2012), Tao *et al.* (2013b) and Kim *et al.* (2016) for large-dimensional Itô processes.

Assumption 2. *We have a sequence T_n of stopping times increasing to infinity, a sequence a_n of bounded positive numbers, such that, for $i = 1, ..., p,\ l = 1, ..., r$,*

$$\begin{aligned} t < T_n \quad \Rightarrow \quad & |Z_t| \leq a_n, \textit{for } Z = \mu_i, \sigma_i^l, \sigma_i^*; \textit{and for } Z = \sigma_i^l, \sigma_i^*, \\ & |E_{\mathcal{F}_{t\wedge T_n}}(Z_{(t+s)\wedge T_n} - Z_{t\wedge T_n})| + E_{\mathcal{F}_{t\wedge T_n}}(Z_{(t+s)\wedge T_n} - Z_{t\wedge T_n})^2 \leq a_n s; \\ & |\sigma_{i,t+s}^l - \sigma_{i,t}^l|^2 + |\sigma_{i,t+s}^* - \sigma_{i,t}^*|^2 \leq a_n s^{1-\epsilon}, \quad \textit{for } \epsilon > 0. \end{aligned}$$

The last regularity condition holds when σ_i^l's and σ_i^*'s follow Brownian Itô process with locally bounded coefficient processes which can be checked by Lévy's continuity theorem. The next assumption gives the condition on the cross-sectional dependence of the specific driving Brownian motions.

Assumption 3. *$k_n/\sqrt{n}$ is bounded from above and from below, $\log p = o(n^{1/2-\epsilon})$ and $\frac{\sqrt{n}}{p^{\delta'}} = o(1)$ for some $\delta' \geq 1$ and any $\epsilon > 0$. $\frac{1}{p}\sum_{i=1}^p \sum_{j=1}^p |\rho_{ij}^*| < M$ for some $M < \infty$.*

The intuition under Assumption 3 is that the specific factors are cross sectionally weakly dependent. It includes the strict factor model as a special case. It is also satisfied when the components of $\boldsymbol{W}^*$ are m-dependent or under some other mixing conditions. Assumption 3 implies that

$$E(\sum_{i=1}^p \frac{F_k^*(i,1)}{\sqrt{p}})^{2\delta'} \leq C(\frac{1}{p}\sum_i \sum_j |\rho_{ij}^*|)^{\delta'} < M,$$

and

$$E\left(\sum_{i=1}^p \frac{(F_k^{*2}(i,1) - 1)}{\sqrt{p}}\right)^{2\delta'} \leq C(\frac{1}{p}\sum_i \sum_j (\rho_{ij}^*)^2)^{\delta'} < M,$$

for all $k = 1, ..., [n/k_n]$.

Next assumption is on the minimum eigenvalue of $\frac{\boldsymbol{\sigma}_t'\boldsymbol{\sigma}_t}{p}$. It is a continuous-time analogue of the factor loading condition imposed in Bai and Ng (2002), and Fan *et al.* (2013) for the discrete approximate factor model.

Assumption 4. *There exists a sequence of stopping times $T_n \uparrow \infty$ and constants $a_n^* > 0$, so that,*

$$\inf_{0 \leq t \leq T_n} \lambda_{min}(\frac{\boldsymbol{\sigma}_t'\boldsymbol{\sigma}_t}{p}) \geq a_n^*, \tag{3.1}$$

and for all $t \in [0, T]$, $\frac{\boldsymbol{\sigma}_t'\boldsymbol{\sigma}_t}{p}$ has distinct eigenvalues.

4. Main Results. We first give the convergence rate of the estimators of the systematic and idiosyncratic risks of an individual Itô process.

THEOREM 1. *Suppose Assumptions 2-4 hold.*

$$\hat{\Gamma}^c_{ii}(\hat{r}) - \int_0^T \boldsymbol{\sigma}'_{is}\boldsymbol{\sigma}_{is}ds = O_p(\frac{1}{C^2_{pn}}),\ \hat{\Sigma}^c_{ii}(s,r) - \Sigma^c_{ii}(s,r) = O_p(\frac{1}{p\wedge n^{1/4}}), \tag{4.1}$$

and

$$\hat{\Gamma}^e_{ii}(\hat{r}) - \int_0^T (\sigma^*_{is})^2 ds = O_p(\frac{1}{C^2_{pn}}),\ \hat{\Sigma}^e_{ii}(s,r) - \Sigma^e_{ii}(s,r) = O_p(\frac{1}{p\wedge n^{1/4}}), \tag{4.2}$$

where $C_{pn} = \sqrt{p}\wedge n^{1/4}$.

REMARK 1. *It is well known in low-dimensional high-frequency context that aggregating $n^{1/4}$-consistent estimates of spot volatilities results in $\sqrt{n}$-consistent estimate of the integrated volatility. An interesting distinctive property in large-dimensional high-frequency data analysis demonstrated in Theorem 1 is that the convergence rate after aggregation achieves $p\wedge\sqrt{n}$ compared with the $p\wedge n^{1/4}$-consistency rate of the spot estimates. If $p \geq C\sqrt{n}$, aggregation yields the consistency rate of $\sqrt{n}$ in contrast to $n^{1/4}$, the rate for the spot analogue, typical as in the low-dimensional setting. However, if $Cn^{1/4} \leq p < C\sqrt{n}$, aggregation results in a rate of p while the rate of the spot estimates is $n^{1/4}$. Even more surprisingly, when $p < Cn^{1/4}$, the integrated systematic or idiosyncratic volatility has the same convergence rate as the local estimate, both being p-consistent.*

Next theorem gives a concentration-type inequality of the maximum of the estimation errors of the residual covariations, which is crucial in proving Theorem 3.

THEOREM 2. *Suppose Assumptions 2-4 hold and $\max_{m\leq p}\frac{1}{\sqrt{p}}\sum_{i=1}^p|\rho^*_{im}| < M$, we have,*

$$P\left(\max_{1\leq i,m\leq p}|\hat{\Gamma}^e_{im}(r) - \Gamma^e_{im}(r)| > h(\frac{\sqrt{\log p}}{n^{1/2}} + \frac{1}{\sqrt{p}})\right) = O(\frac{n^{1/2}}{p^{\delta'}} + p^{\frac{-\delta'}{2}} + p^{1-\delta'}n^{1-\delta'/2}), \tag{4.3}$$

and

$$P\left(\max_{1\leq i,m\leq p}|\hat{\boldsymbol{\Sigma}}^e_{im}(s,r) - \boldsymbol{\Sigma}^e_{im}(s,r)| > h(\frac{\sqrt{\log p}}{n^{1/4}} + \frac{1}{\sqrt{p}})\right) = O(\frac{n^{1/2}}{p^{\delta'}} + p^{\frac{-\delta'}{2}} + p^{1-\delta'}n^{1-\delta'/2}) \tag{4.4}$$

for some constant $h > 0$.

Remark 2. *Theorem 2 demonstrates that the estimated idiosyncratic covolatilities are uniformly close to the realized ones. This makes identifying the "signals" using estimated idiosyncratic covolatilities feasible. Similar to Remark 1, aggregation leads to rate enhancement when $n < Cp\log p$.*

Theorem 3 below reveals the convergence rate of the thresholding estimator of the sparse idiosyncratic volatility matrices.

Theorem 3. *Suppose Assumptions 1-4 hold, $\max_{m\le p}\frac{1}{\sqrt{p}}\sum_{i=1}^{p}|\rho^*_{im}| < M$, and $\lambda_{max}(\boldsymbol{\rho}^*) < c$ for some constant $c > 0$. We have,*

$$\begin{aligned} & P\left(\sup_{\boldsymbol{\rho}^*\in\mathcal{U}_q(s_0(p))}\|\hat{\boldsymbol{\Gamma}}^{e\mathcal{T}}(\hat{r})-\boldsymbol{\Gamma}^e\|\le C_q s_0(p)\lambda_{pn}^{1-q}\right)\\ = & \ 1-O(p^{-\delta'}\sqrt{n}+p^{-\delta'/2}+p^{1-\delta'}n^{1-\delta'/2})+o(1), \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} & P\left(\sup_{\boldsymbol{\rho}^*\in\mathcal{U}_q(s_0(p))}\|\hat{\boldsymbol{\Sigma}}^{e\mathcal{T}}(s,\hat{r})-\boldsymbol{\Sigma}^e(s)\|\le C_q s_0(p)\tilde{\lambda}_{pn}^{1-q}\right)\\ = & \ 1-O(p^{-\delta'}\sqrt{n}+p^{-\delta'/2}+p^{1-\delta'}n^{1-\delta'/2})+o(1), \end{aligned} \tag{4.6}$$

for some constant C_q, where $\lambda_{pn}=\frac{1}{\sqrt{p}}+\frac{\sqrt{\log p}}{\sqrt{n}}$ and $\tilde{\lambda}_{pn}=\frac{1}{\sqrt{p}}+\frac{\sqrt{\log p}}{n^{1/4}}$.

Remark 3. *Theorem 3 demonstrates that the convergence rate of the thresholding estimator of the sparse integrated idiosyncratic volatility matrix is of order $s_0(p)\lambda_{pn}^{1-q}$. If $s_0(p)\lambda_{pn}^{1-q}=o(1)$ and $p^{-\delta'}\sqrt{n}+p^{1-\delta'}n^{1-\delta'/2}=o(1)$, $\hat{\boldsymbol{\Gamma}}^{e\mathcal{T}}(\hat{r})$ is a consistent estimator in terms of the spectral norm. Compared with the thresholding level used in Bickel and Levina (2008) and Cai and Liu (2011), the level used in the present paper has an additional term $\frac{1}{\sqrt{p}}$. This comes from the estimation error of the factor loadings and the factor scores. So in thresholding the realized integrated idiosyncratic volatility matrix, an estimated "signal" could be identified only if it has a strength stronger than the combined estimation error caused in estimating the latent residual process and in estimating the covariations of the error processes. The $o(1)$ term comes from the estimation of the number of common factors. Similar conclusions can be drawn on the estimated spot idiosyncratic volatility matrix.*

A straight corollary of Theorem 3 and the identity

$$(\hat{\boldsymbol{\Gamma}}^{e\mathcal{T}}(\hat{r}))^{-1}-(\boldsymbol{\Gamma}^e)^{-1}=(\hat{\boldsymbol{\Gamma}}^{e\mathcal{T}}(\hat{r}))^{-1}(\boldsymbol{\Gamma}^e-\hat{\boldsymbol{\Gamma}}^{e\mathcal{T}}(\hat{r}))(\boldsymbol{\Gamma}^e)^{-1}$$

is the following.

Corollary 1. *Under the conditions in Theorem 3, if* $s_0(p)\lambda_{pn}^{1-q}=o(1)$, $p^{-\delta'}\sqrt{n}+p^{1-\delta'}n^{1-\delta'/2}=o(1)$, $\inf_{0\le s\le T}\min_{1\le i\le p}(\sigma_{is}^*)^2>c^{-1}$ *and* $c^{-1}\le\lambda_{min}(\boldsymbol{\rho}^*)\le\lambda_{max}(\boldsymbol{\rho}^*)<c$ *for some constant* $c>0$*, we have*

$$\|(\hat{\boldsymbol{\Gamma}}^{e\mathcal{T}}(\hat r))^{-1}-(\boldsymbol{\Gamma}^e)^{-1}\|=O_p(s_0(p)\lambda_{pn}^{1-q}), \tag{4.7}$$

and

$$\|(\hat{\boldsymbol{\Sigma}}^{e\mathcal{T}}(s,\hat r))^{-1}-(\boldsymbol{\Sigma}^e(s))^{-1}\|=O_p(s_0(p)\tilde\lambda_{pn}^{1-q}). \tag{4.8}$$

Now we give a result on the factor-based estimator of the inverse of the spot volatility matrix.

Theorem 4. *Under the conditions in Corollary 1, we have*

$$\|\hat{\boldsymbol{\Sigma}}_s^{-1}(\hat r)-\boldsymbol{\Sigma}_s^{-1}\|=O_p(\tilde\lambda_{pn}^{1-q}s_0(p)+(\frac{1}{\sqrt{p}}+n^{-1/4})+\frac{1}{n^{1/8}p^{3/8}}). \tag{4.9}$$

Remark 4. *In Theorem 4,* $\tilde\lambda_{pn}^{1-q}s_0(p)$ *is due to the error accumulation in estimating the idiosyncratic volatility matrix,* $\boldsymbol{\Sigma}_s^e$*. The term* $\frac{1}{\sqrt{p}}+\frac{1}{n^{1/8}p^{3/8}}$ *is a result of the cross-section dependence of the idiosyncratic factors, while the term* $n^{-1/4}$ *comes from the discretization error. Therefore, as discussed in the introduction, if the holding period of certain portfolio is short,* $\hat{\boldsymbol{\Sigma}}_s^{-1}(\hat r)$ *serves as a good tool in portfolio allocation. The discretization error term does not appear in Fan et al. (2013) for discrete approximate factor model and Aït-Sahalia and Xiu (2016b) for continuous-time factor model since constant loadings are assumed. As in all the factor-based precision matrix estimator, the proof of this theorem depends heavily on the Sherman-Morrison-Woodbury formula, which, however, does not carry through for the inverse of the factor-based integrated volatility matrix via aggregation,* $\hat{\boldsymbol{\Gamma}}(\hat r)\equiv\hat{\boldsymbol{\Gamma}}^c(\hat r)+\hat{\boldsymbol{\Gamma}}^{e\mathcal{T}}(\hat r)$*. To be clearer, let*

$$\hat{\boldsymbol{\Theta}}=\sqrt{k_n\Delta_n}(\hat{\boldsymbol{\sigma}}_1,...,\hat{\boldsymbol{\sigma}}_{[n/k_n]})\ \text{and}\ \tilde{\boldsymbol{\Theta}}=\sqrt{k_n\Delta_n}(\boldsymbol{\sigma}_1\boldsymbol{H}_1^{-1},...,\boldsymbol{\sigma}_{[n/k_n]}\boldsymbol{H}_{[n/k_n]}^{-1}).$$

Then by definition, we have, in matrix form, $\hat{\boldsymbol{\Gamma}}^c=\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\Theta}}'$*. Let* $\tilde{\boldsymbol{\Gamma}}^c=\tilde{\boldsymbol{\Theta}}\tilde{\boldsymbol{\Theta}}'$ *and* $\tilde{\boldsymbol{\Gamma}}=\tilde{\boldsymbol{\Gamma}}^c+\boldsymbol{\Gamma}^e$*. Then we have* $\|\hat{\boldsymbol{\Gamma}}^{-1}(\hat r)-\boldsymbol{\Gamma}^{-1}\|\le\|\hat{\boldsymbol{\Gamma}}^{-1}(\hat r)-\tilde{\boldsymbol{\Gamma}}^{-1}\|+\|\tilde{\boldsymbol{\Gamma}}^{-1}-\boldsymbol{\Gamma}^{-1}\|$*. To bound the first term, following the proof of Theorem 4, in (6.33), one wishes* $\lambda_{min}(\tilde{\boldsymbol{\Theta}}'\tilde{\boldsymbol{\Theta}})\ge Cp$ *which is not true whenever* $p\ne[n/k_n]r$*. So estimating the inverse of the integrated volatility matrix under model (1.1) via aggregating local factor analysis seems infeasible, and we leave it as a future problem. Under model (1.2), however, this becomes simple and global PCA (a special case of local PCA when* $k_n=n$*) as in Aït-Sahalia and Xiu (2016b) works since the discretization error term in Theorem 4 is vanishing and local analysis becomes unnecessary.*

5. Numerical Studies.

5.1. *Simulation Studies.* In this section, we conduct simulations to check the performance of our estimators. We generate data from the stochastic volatility models of the form (1.1) with $r = 3$. The instantaneous volatility processes, σ_{it}^l's are generated independently from the following square-root processes,

$$(\sigma_{it}^l)^2 = b_{li}(a_{li} - (\sigma_{it}^l)^2)dt + \sigma_{li}^0 \sigma_{it}^l dW_{it}^\sigma,\ l = 1, ..., r.$$

We set $a_{1i} = 0.5 + i/p$, $a_{2i} = 0.75 + i/p$, $a_{3i} = 0.6 + i/p$, $b_{1i} = 0.03 + i/100p$, $b_{2i} = 0.05 + i/100p$, $b_{3i} = 0.08 + i/100p$, $\sigma_{1i}^0 = 0.15 + i/10p$, $\sigma_{2i}^0 = \sigma_{3i}^0 = 0.2 + i/10p$. These parameters are similarly chosen as in Jacod and Todorov (2014) and Kong *et al.* (2015). The specific volatility process follows the stochastic differential equation,

$$(\sigma_{it}^*)^2 = (0.08 + i/100p)(0.25 + i/p - (\sigma_{it}^*)^2 dt + (0.2 + i/10p)\sigma_{it}^* dW_{it}^{\sigma*}.$$

The simulations are repeated for 2000 times. We set $p = 100, 200$. As in Kong (2016a), k_m is set to be 6 and $\beta = \hat{\zeta} = U(6, \tilde{\boldsymbol{F}}^6)$, and the penalty function used in determining r is $g(p, n) = \frac{p+k_n}{pk_n} \log \frac{pk_n}{p+k_n}$.

We first consider the case when $n = 1170$ mimicking every-one-minute data set of 3 days (3×390). We let $k_n = 30 \approx \sqrt{1170}$ and hence the data set is split into 39 blocks. The simulation results are reported in the upper half of Table 5.1, which displays the averaged squared relative estimation errors of the **I**ntegrated **S**ystematic **V**olatility (EISV) and **I**diosyncratic **V**olatility (EIIV) of the first simulated process, and averaged absolute estimation error of the **T**hresholded **I**ntegrated **I**diosyncratic **V**olatility **M**atrix (ETIIVM). EIIVM stands for the averaged absolute estimation error of the realized integrated volatility matrix without thresholding. In estimating the integrated idiosyncratic volatility matrix, we set the thresholding parameter, h, so that the averaged squared spectral norm of the estimation error stops trending lower.

TABLE 1

Averaged squared relative estimation errors of the integrated systematic volatility (EISV) and idiosyncratic volatility (EIIV) of the first simulated process, and averaged absolute estimation error in spectral norm of the sparse integrated idiosyncratic volatility matrix. EIIVM stands for the averaged absolute estimation error in spectral norm of the realized integrated volatility matrix without thresholding, and EIIIVM for that of the thresholding estimator.

		$n = 1170$	3×390	
	EISV	EIIV	EIIVM	ETIIVM
$p = 100$	0.0017	0.0075	0.1882	0.0671
$p = 200$	0.0019	0.0072	0.1887	0.0589
		$n = 780$	10×78	
	EISV	EIIV	EIIVM	ETIIVM
$p = 100$	0.0029	0.0131	0.6749	0.2551
$p = 200$	0.0027	0.0086	0.6849	0.2377

Next, we consider the case when $n = 780$ mimicking every-five-minute data set of 10 days (10×78). We let $k_n = 30 \approx \sqrt{780}$ and hence the data set is split into 26 blocks. The simulation results are reported in the lower half of Table 5.1. From the table, we observe that our estimators are quite close to the true parameters across the board. The thresholding estimator of the integrated idiosyncratic volatility matrix performs better than the one without thresholding. This is consistent to Theorem 3 and the well known fact that the realized integrated idiosyncratic volatility matrix works bad when p is large.

5.2. *Real Data Analysis.* In this section, we implement our estimators to a real financial data set consisting of 99 heavily traded stocks included in S&P 500 index in April 2013. We start from April 1st and remove the transaction prices within the first 20 minutes of April 1st after the opening of the market in order to get rid of the cross-month jump effect. In order to avoid the adverse effect of microstrcture noise, we sparsely sample the data set. We choose to use $n = 780$ observations of 9-minute log returns. We set $k_n = 30$ and estimate the number of r by using the same settings as in Kong (2016a) which results in $\hat{r} = 5$. The largest eigenvalue of the factor-based estimated integrated idiosyncratic volatility matrix is $1.0e - 004 * 0.0025$. To show robustness of the estimation over time, we remove the transaction prices within the first 30 minutes of April 01 after the opening of the market and calculate the estimators using $n = 780$ observations of 9-minute log returns. This time, the largest eigenvalue of the factor-based estimated integrated idiosyncratic volatility matrix is $1.0e - 004 * 0.0020$. Comparing these two cases, we see that the estimated realized largest eigenvalues are quite close since the second data set is only 10 minutes ahead of the first one.

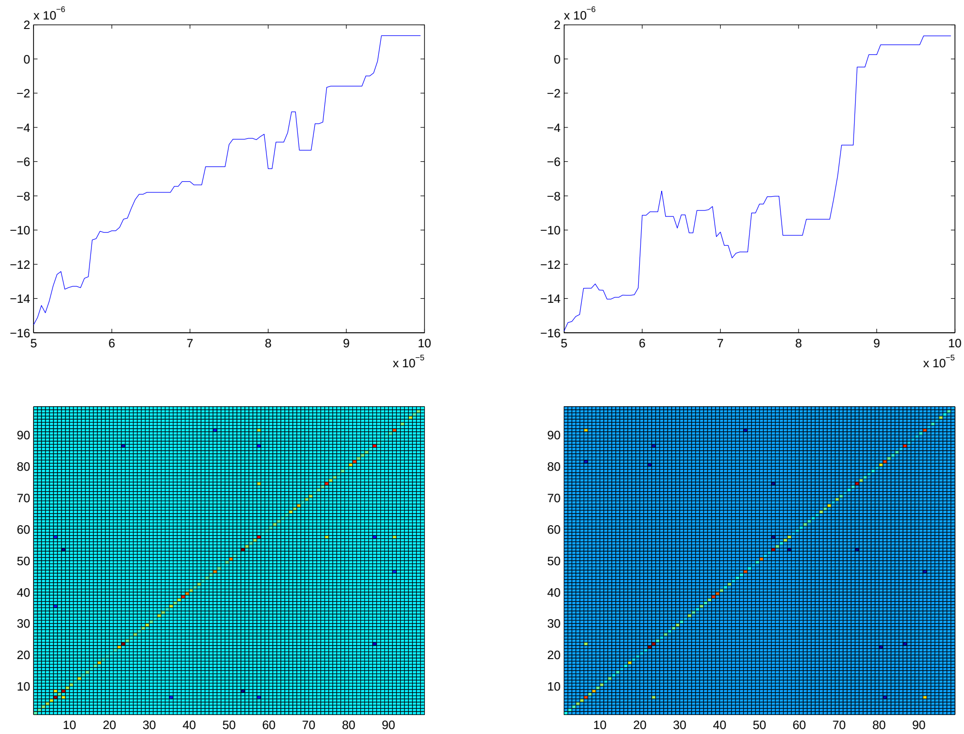
In thresholding, we choose the tuning parameter as in Fan *et al.* (2013), that is, we choose h to be the smallest real number in a fine grid so that $\hat{\mathbf{\Gamma}}^{e\mathcal{T}}(\hat{r})$ is positive definite. The upper panels in Figure 1 display the minimum eigenvalues of $\hat{\mathbf{\Gamma}}^{e\mathcal{T}}(\hat{r})$ as a function of h. Left panel corresponds the first data set and the right panel the second data set. Both panels show that the chosen h is around $9 \times 1.0e - 005$.

The lower panels in Figure 1 display the heat maps of the thresholding estimator of the integrated idiosyncratic volatility matrix. From the figure, there is strong evidence of sparsity structure. It also demonstrates that the specific factors are not independent.

6. Conclusion and Discussion. In this paper, we proposed estimators of the (integrated and spot) systematic and idiosyncratic volatilities and also estimators of the sparse (integrated and spot) idiosyncratic volatility matrix and the inverse of the spot volatility matrix. There are several interesting problems related to this topic that are either undergoing or to be studied in the future.

- Testing for the bandedness of the integrated idiosyncratic volatility matrix is of importance in econometrics. For discrete-time models, we refer to Qiu and Chen (2012).
- There are some practical issues to be considered when using ultra-high- frequency data. How large is the bias caused by the asynchronicity and mi-

FIG 1. *Upper Panels: The minimum eigenvalues of* $\hat{\mathbf{\Gamma}}^{e\mathcal{T}}(\hat{r})$ *as a function of* h*; Lower Panels: Heat maps of the thresholding estimator of the integrated idiosyncratic volatility matrix; Left panels: The first data set; Right panels: The second data set.*



crostructure noise and how to get rid of it? There are many efficient noise smoothing methods, like the two-time scale technique in Zhang *et al.* (2005) and the pre-averaging approach invented in Jacod *et al.* (2009). We conjecture that a hybrid of our estimation procedure and any noise smoothing technique would result in noise-robust estimators with a sacrificed convergence rate (replace n in Theorems 1-4 to $\sqrt{n}$.) We refer to Wang and Zou (2010), Tao *et al.* (2013a) (2013b), Kim *et al.* (2015) for large volatility matrix estimation with measurement errors (but without factor structure).

Appendix A: Proof of the Main Theorems

In the proof, C stands for a generic constant that may take different values at different appearances. $\mathcal{F}_{k-1} = \mathcal{F}_{t_{(k-1)k_n\Delta_n}}$. By the standard localization method, it suffices to prove the results under the following strengthened assumption.

Assumption 5. *For $l=1,...,r$,*

$$\max_{1\le i\le p}\sup_{0\le t\le T}(|\sigma_{it}^l|+|\sigma_{it}^*|+|\mu_{it}|)<C,\ \inf_{0\le t\le T}\lambda_{min}(\frac{\boldsymbol{\sigma}_t'\boldsymbol{\sigma}_t}{p})\ge C^{-1}, \tag{6.1}$$

$$\text{and}\quad \inf_{0\le t\le T}\min_{1\le l\le r-1}|\lambda_{l+1}(\frac{\boldsymbol{\sigma}_t'\boldsymbol{\sigma}_t}{p})-\lambda_l(\frac{\boldsymbol{\sigma}_t'\boldsymbol{\sigma}_t}{p})|\ge C^{-1},$$

$$\max_{1\le i\le p}[E_t(Z_{t+s}-Z_t)^2+|E_t(Z_{t+s}-Z_t)|\le Cs,\ Z=\sigma_i^l,\sigma_i^*,$$

$$|\sigma_{i,t+s}^l-\sigma_{i,t}^l|^2+|\sigma_{i,t+s}^*-\sigma_{i,t}^*|^2\le Cs^{1-\epsilon},\quad \text{for } \epsilon>0. \tag{6.2}$$

By (2.5), it is enough to prove the theorems by assuming that $r\le r_m$ is known. We will assume this in the sequel.

Important decompositions Let $\boldsymbol{V}_k$ be $r\times r$ diagonal matrix of the r largest eigenvalues of $\frac{\boldsymbol{\delta}_k'\boldsymbol{\delta}_k}{pk_n}$ in decreasing order. Then we have $\frac{1}{pk_n}\boldsymbol{\delta}_k'\boldsymbol{\delta}_k\hat{\boldsymbol{F}}_k'=\hat{\boldsymbol{F}}_k'\boldsymbol{V}_k$. By this and the decomposition for $\boldsymbol{\delta}_k$, we soon have the following decomposition,

$$\hat{\boldsymbol{F}}_k(j)-\boldsymbol{H}_k\boldsymbol{F}_k(j)=I_{kj}+II_{kj}+III_{kj}+IV_{kj}+V_{kj}, \tag{6.3}$$

where

$$\begin{aligned}
I_{kj} &= \boldsymbol{V}_k^{-1}\frac{1}{k_n}\sum_{l=1}^{k_n}\hat{\boldsymbol{F}}_k(l)\frac{1}{p}\sum_{i=1}^{p}\sigma_{ki}^{*2}\left(F_k^*(i,l)F_k^*(i,j)-E[F_k^*(i,l)F_k^*(i,j)]\right)\\
II_{kj} &= \boldsymbol{V}_k^{-1}\frac{1}{k_n}\sum_{l=1}^{k_n}\hat{\boldsymbol{F}}_k(l)\frac{1}{p}\sum_{i=1}^{p}\sigma_{ki}^{*2}E\left[F_k^*(i,l)F_k^*(i,j)\right]\\
III_{kj} &= \boldsymbol{V}_k^{-1}\frac{1}{k_n}\sum_{l=1}^{k_n}\hat{\boldsymbol{F}}_k(l)\left(\frac{1}{p}\boldsymbol{F}_k'(l)\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*(j)+\frac{1}{p}\boldsymbol{F}_k^{*\prime}(l)\boldsymbol{\sigma}_k^{*\prime}\boldsymbol{\sigma}_k\boldsymbol{F}_k(j)\right)\\
IV_{kj} &= \boldsymbol{V}_k^{-1}\frac{1}{k_n}\sum_{l=1}^{k_n}\hat{\boldsymbol{F}}_k(l)\frac{1}{p}\left(\boldsymbol{\delta}_k'(l)-\overline{\boldsymbol{\delta}}_k'(l)\right)\boldsymbol{\delta}_k(j)\\
V_{kj} &= \boldsymbol{V}_k^{-1}\frac{1}{k_n}\sum_{l=1}^{k_n}\hat{\boldsymbol{F}}_k(l)\frac{1}{p}\overline{\boldsymbol{\delta}}_k'(l)\left(\boldsymbol{\delta}_k(j)-\overline{\boldsymbol{\delta}}_k(j)\right),
\end{aligned}$$

and $\boldsymbol{A}(j)$ stands for the jth column of the matrix $\boldsymbol{A}$, $A(i,l)$ stands for the entry in row i and column l of the matrix $\boldsymbol{A}$, and $\sigma_{ki}^*=\boldsymbol{\sigma}_k^*(i,i)$. In matrix form,

$$\begin{aligned}
\hat{\boldsymbol{F}}_k-\boldsymbol{H}_k\boldsymbol{F}_k &= \frac{\boldsymbol{V}_k^{-1}\hat{\boldsymbol{F}}_k}{pk_n}\Big((\boldsymbol{\delta}_k-\bar{\boldsymbol{\delta}}_k)'\boldsymbol{\delta}_k+\bar{\boldsymbol{\delta}}_k'(\boldsymbol{\delta}_k-\bar{\boldsymbol{\delta}}_k)\\
&\quad +\boldsymbol{F}_k^{*\prime}\boldsymbol{\sigma}_k^{*\prime}\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*+\boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*+\boldsymbol{F}_k^{*\prime}\boldsymbol{\sigma}_k^*\boldsymbol{\sigma}_k\boldsymbol{F}_k\Big).
\end{aligned} \tag{6.4}$$

Recall that $\hat{\boldsymbol{\sigma}}_k=\boldsymbol{\delta}_k\hat{\boldsymbol{F}}_k'/k_n$, then by the fact that $\hat{\boldsymbol{F}}_k\hat{\boldsymbol{F}}_k'/k_n=I_r$, we have

$$\hat{\boldsymbol{\sigma}}_k=\boldsymbol{\sigma}_k\boldsymbol{H}_k^{-1}+VI_k+VII_k, \tag{6.5}$$

where

$$
\begin{aligned}
VI_k &= \frac{1}{k_n}\boldsymbol{\sigma}_k\left(\boldsymbol{F}_k - \boldsymbol{H}_k^{-1}\hat{\boldsymbol{F}}_k\right)\hat{\boldsymbol{F}}_k' \\
VII_k &= \frac{1}{k_n}\left(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k + \boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\right)\hat{\boldsymbol{F}}_k'.
\end{aligned}
$$

Then we soon have

$$
\hat{\boldsymbol{\sigma}}_{ki} = \boldsymbol{H}_k^{-1\prime}\boldsymbol{\sigma}_{ki} + VI_{ki} + VII_{ki}, \tag{6.6}
$$

where $\boldsymbol{A}_{ki}'$ always denotes the ith row of $\boldsymbol{A}_k = (\boldsymbol{A}_{k1}, ..., \boldsymbol{A}_{kl})'$, $l = r, k_n$. Combining (6.3) and (6.6), we have

$$
\hat{\boldsymbol{\sigma}}_{ki}'\hat{\boldsymbol{F}}_k(j) = \boldsymbol{\sigma}_{ki}'\boldsymbol{F}_k(j) + R_{ij}^k, \tag{6.7}
$$

where

$$
\begin{aligned}
R_{ij}^k &= \boldsymbol{\sigma}_{ki}'\boldsymbol{H}_k^{-1}(I_{kj} + II_{kj} + III_{kj} + IV_{kj} + V_{kj}) \\
&\quad + (VI_{ki}' + VII_{ki}')(\boldsymbol{H}_k\boldsymbol{F}_k(j) + I_{kj} + II_{kj} + III_{kj} + IV_{kj} + V_{kj}).
\end{aligned}
$$

From (6.7), one easily gets

$$
\frac{1}{k_n}\sum_{j=1}^{k_n}(\hat{\boldsymbol{\sigma}}_{ki}'\hat{\boldsymbol{F}}_k(j))^2 = \frac{1}{k_n}\sum_{j=1}^{k_n}(\boldsymbol{\sigma}_{ki}'\boldsymbol{F}_k(j))^2 + \frac{1}{k_n}\sum_{j=1}^{k_n}(R_{ij}^k)^2 + \frac{2}{k_n}\sum_{j=1}^{k_n}\boldsymbol{\sigma}_{ki}'\boldsymbol{F}_k(j)R_{ij}^k. \tag{6.8}
$$

Before stating the proof of Theorem 1, we introduce some lemmas, of which the proofs are given either in Appendix B or the supplement.

Lemma 1. *Let $\boldsymbol{V}_k^0$ and $\boldsymbol{\gamma}_k^0$ be the eigenvalue matrix and eigenvector matrix of $\boldsymbol{B}_k^0 \equiv \frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p}$, respectively. Under Assumptions 2-4, if $p = e^{o(k_n^{1-\epsilon})}$ for any $\epsilon > 0$, we have for any $M > 0$ and some $h > 0$,*

$$
P\left(\|\boldsymbol{V}_k - \boldsymbol{V}_k^0\| > h(\frac{1}{p^{1/4}} + \frac{\sqrt{\log p}}{k_n^{1/2-\epsilon}})\right) = O(p^{-\frac{\delta'}{2}\vee(-M)}), \tag{6.9}
$$

$$
P\left(\|\boldsymbol{H}_k - (\boldsymbol{V}_k^0)^{-1/2}\boldsymbol{\gamma}_k^{0\prime}(\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}\| > h(\frac{1}{p^{1/4}} + \sqrt{\frac{\log p}{k_n^{1-\epsilon}}})\right) = O(p^{-\frac{\delta'}{2}\vee(-M)}), \tag{6.10}
$$

$$
\text{and } P\left(\|\boldsymbol{H}_k'\boldsymbol{H}_k - I_r\| > h(\frac{1}{p^{1/4}} + \sqrt{\frac{\log p}{k_n^{1-\epsilon}}})\right) = O(p^{-\frac{\delta'}{2}\vee(-M)}). \tag{6.11}
$$

Corollary 2. *If $p^{-\delta'} k_n = o(1)$ and $p = e^{o(k_n^{1-\epsilon})}$, we have for some $h' > 0$*

$$P\left(\max_k(\|\boldsymbol{V}_k\| + \|\boldsymbol{V}_k^{-1}\| + \|\boldsymbol{H}_k\| + \|\boldsymbol{H}_k^{-1}\| + \frac{\|\hat{\boldsymbol{F}}_k - \boldsymbol{H}_k\boldsymbol{F}_k\|}{\sqrt{k_n}}) > h'\right)$$

$$= O(p^{-\delta'} k_n) = o(1). \tag{6.12}$$

Lemma 2. *Let $d > 0$ be an integer, under Assumptions 2-4, we have*

$$\max_{1\le k\le [n/k_n]} \|\frac{1}{k_n}\sum_{j=1}^{k_n} \boldsymbol{F}_k(j)\boldsymbol{F}_k'(j)\| = O_p(1), \tag{6.13}$$

$$E_{\mathcal{F}_{k-1}}\left(\|\boldsymbol{V}_k^{-1} - (\boldsymbol{V}_k^0)^{-1}\| + \|\boldsymbol{V}_k - \boldsymbol{V}_k^0\| + \|\boldsymbol{H}_k^{-1} - (\boldsymbol{H}_k^0)^{-1}\| + \|\boldsymbol{H}_k - \boldsymbol{H}_k^0\|\right)^d \le C(\frac{1}{C_{pn}})^d \tag{6.14}$$

$$E_{\mathcal{F}_{k-1}}\left(\|\boldsymbol{V}_k^{-1}\| + \|\boldsymbol{V}_k\| + \|\boldsymbol{H}_k^{-1}\| + \|\boldsymbol{H}_k\|\right)^d \le C; \tag{6.15}$$

$$E_{\mathcal{F}_{k-1}}\left(\frac{1}{k_n}\sum_{l=1}^{k_n}\|\hat{\boldsymbol{F}}_k(l) - \boldsymbol{H}_k\boldsymbol{F}_k(l)\|^2\right)^d \le \frac{C}{C_{pn}^{2d}}. \tag{6.16}$$

Let $R_{ij}^k = R_{ij}^k(1) + R_{ij}^k(2)$ where

$$R_{ij}^k(1) = \boldsymbol{\sigma}_{ki}'\boldsymbol{H}_k^{-1}(I_{kj} + II_{kj} + III_{kj} + IV_{kj} + V_{kj}) + (VI_{ki}' + VII_{ki}')\boldsymbol{H}_k\boldsymbol{F}_k(j)$$

$$R_{ij}^k(2) = (VI_{ki}' + VII_{ki}')(I_{kj} + II_{kj} + III_{kj} + IV_{kj} + V_{kj}).$$

Lemma 3. *Under Assumptions 2-4, we have*

$$\sum_{k=1}^{[n/k_n]} \Delta_n \sum_{j=1}^{k_n} \boldsymbol{\sigma}_{ki}'\boldsymbol{F}_k(j)R_{ij}^k(1) = O_p(\frac{1}{C_{pn}^2}). \tag{6.17}$$

$$\sum_{k=1}^{[n/k_n]} \Delta_n \sum_{j=1}^{k_n} \boldsymbol{\sigma}_{ki}'\boldsymbol{F}_k(j)R_{ij}^k(2) = O_p(\frac{1}{C_{pn}^2}). \tag{6.18}$$

Proof of Theorem 1 By the definition of R_{ij}^k, we have

$$(R_{ij}^k)^2 \le C(\|\boldsymbol{\sigma}_{ki}'\boldsymbol{H}_k^{-1}\|^2\|\hat{\boldsymbol{F}}_k(j) - \boldsymbol{H}_k\boldsymbol{F}_k(j)\|^2 + \|\boldsymbol{H}_k\boldsymbol{F}_k(j)\|^2\|\hat{\boldsymbol{\sigma}}_{ki} - \boldsymbol{H}_k^{-1\prime}\boldsymbol{\sigma}_{ki}\|^2). \tag{6.19}$$

Then by Lemma 2, (S.2.46) in the supplement and Corollary 2, we have

$$\sum_{k=1}^{[n/k_n]} \Delta_n \sum_{j=1}^{k_n} (R_{ij}^k)^2 = O_p(\frac{1}{C_{pn}^2}).$$

This together with Lemma 3, (6.8), and the fact that

$$\sum_{k=1}^{[n/k_n]}\sum_{j=1}^{k_n}(\boldsymbol{\sigma}'_{ki}\boldsymbol{F}_k(j))^2=\int_0^T \boldsymbol{\sigma}'_{is}\boldsymbol{\sigma}_{is}ds+O_p(\sqrt{\Delta_n}),$$

proves the first equation of (4.1). Again, by Lemma 2, (S.2.46) in the supplement and Corollary 2, plus (6.19), we have

$$\frac{1}{k_n}\sum_{j=1}^{k_n}(R_{ij}^k)^2=O_p(\frac{1}{C_{pn}^2}).$$

Following the proof of Lemma 3 in the supplement, one easily gets

$$\frac{1}{k_n}\sum_{j=1}^{k_n}\boldsymbol{\sigma}'_{ki}\boldsymbol{F}_k(j)R_{ij}^k(1)=O_p(\frac{1}{p\wedge\sqrt{k_n}}),\ \text{and}\ \frac{1}{k_n}\sum_{j=1}^{k_n}\boldsymbol{\sigma}'_{ki}\boldsymbol{F}_k(j)R_{ij}^k(2)=O_p(\frac{1}{p\wedge\sqrt{k_n}}).$$

The above two equations and Assumption 2 prove the second equation of (4.1). Notice that $\int_0^t\boldsymbol{\sigma}'_{is}\boldsymbol{\sigma}_{is}ds+\int_0^t(\sigma_{is}^*)^2ds$ and $\boldsymbol{\sigma}'_{is}\boldsymbol{\sigma}_{is}+(\boldsymbol{\sigma}_{is}^*)^2$ are the integrated volatility and spot volatility of X_i, respectively. On the other hand, $\sum_{k=1}^{[n/k_n]}\Delta_n\sum_{j=1}^{k_n}(\delta_{ij}^k)^2$ and $\frac{1}{k_n}\sum_{j=1}^{k_n}(\delta_{ij}^s)^2$ are realized integrated and local variances, respectively. Then, it is well known that they converge to the integrated volatility and spot volatility at the rate of $O_p(\sqrt{\Delta_n})$ and $O_p(\Delta_n^{1/4})$, respectively. This together with (4.1) proves (4.2).

Proof of Theorem 2 We first get a decomposition on $\hat{\Gamma}_{im}^e-\Gamma_{im}^e$. By (6.7), we have

$$\delta_k(i,j)-\hat{\boldsymbol{\sigma}}'_{ki}\hat{\boldsymbol{F}}_k(j)=\delta_k(i,j)-\overline{\delta}_k(i,j)+\sigma_{ki}^*F_k^*(i,j)-R_{ij}^k. \tag{6.20}$$

Then we further have

$$\begin{aligned}&(\delta_k(i,j)-\hat{\boldsymbol{\sigma}}'_{ki}\hat{\boldsymbol{F}}_k(j))(\delta_k(m,j)-\hat{\boldsymbol{\sigma}}'_{km}\hat{\boldsymbol{F}}_k(j))\\=\ &\sigma_{ki}^*F_k^*(i,j)F_k^*(m,j)\sigma_{km}^*+(\delta_k(i,j)-\overline{\delta}_k(i,j)-R_{ij}^k)\\&\times(\delta_k(m,j)-\overline{\delta}_k(m,j)-R_{mj}^k)\\&+\sigma_{ki}^*F_k^*(i,j)(\delta_k(m,j)-\overline{\delta}_k(m,j)-R_{mj}^k)\\&+\sigma_{km}^*F_k^*(m,j)(\delta_k(i,j)-\overline{\delta}_k(i,j)-R_{ij}^k).\end{aligned}\tag{6.21}$$

By the definition of $\hat{\Gamma}_{im}^e$ and (6.21), it suffices to prove the following lemmas of which the proofs are given in the supplement. Lemmas 4 and 5 are used, respectively, to prove (4.3) and (4.4).

Lemma 4. *For some $h>0$, we have under assumptions in Theorem 2,*

$$P\left(\max_i \sum_{k=1}^{[n/k_n]} \Delta_n \sum_{j=1}^{k_n} (\delta_k(i,j)-\overline{\delta}_k(i,j)-R^k_{ij})^2 > h(\frac{\sqrt{\log p}}{k_n}+\frac{1}{\sqrt{p}})\right)$$
$$= O(p^{-\delta'/2}+p^{-\delta'}k_n), \quad (6.22)$$

and
$$P\left(\max_{i,m}|\sum_{k=1}^{[n/k_n]} \Delta_n \sum_{j=1}^{k_n} \sigma^*_{ki}F^*_k(i,j)(\delta_k(m,j)-\overline{\delta}_k(m,j)-R^k_{mj})|\right.$$
$$\left. > h(\frac{\sqrt{\log p}}{k_n}+\frac{1}{\sqrt{p}})\right) = O(p^{-\delta'/2}+p^{-\delta'}k_n+p^{1-\delta'}k_n^{2-\delta'}). \quad (6.23)$$

Lemma 5. *For some $h>0$, we have under assumptions in Theorem 2,*

$$P\left(\max_i \frac{1}{k_n}\sum_{j=1}^{k_n} (\delta_s(i,j)-\overline{\delta}_s(i,j)-R^s_{ij})^2 > h(\frac{\sqrt{\log p}}{\sqrt{k_n}}+\frac{1}{\sqrt{p}})\right)$$
$$= O(p^{-\delta'/2}+p^{-\delta'}k_n), \quad (6.24)$$

and
$$P\left(\max_{i,m}|\frac{1}{k_n}\sum_{j=1}^{k_n} \sigma^*_{is}F^*_s(i,j)(\delta_s(m,j)-\overline{\delta}_s(m,j)-R^s_{mj})|\right.$$
$$\left. > h(\frac{\sqrt{\log p}}{\sqrt{k_n}}+\frac{1}{\sqrt{p}})\right) = O(p^{-\delta'/2}+p^{-\delta'}k_n+p^{1-\delta'}k_n^{2-\delta'}), \quad (6.25)$$

where R^s_{ij} is similarly defined as R^k_{ij} except for replacing the data starting point $t_{(k-1)k_n}$ by $\lceil s/\Delta_n\rceil\Delta_n$.

Proof of Theorem 3 Let $\mathcal{A}=\{\max_{1\le i\le p}\max_{1\le m\le p}|\hat{\Gamma}^e_{im}(r)-\Gamma^e_{im}|\le h(\frac{1}{\sqrt{p}}+\frac{\sqrt{\log p}}{k_n})\}$. Then on $\mathcal{A}$, we have by setting $\lambda=\lambda_{pn}=h(\frac{1}{\sqrt{p}}+\frac{\sqrt{\log p}}{k_n})$, Assumption 5 and the property of $g_\lambda(z)$,

$$\|\hat{\mathbf{\Gamma}}^{e\mathcal{T}}(r)-\mathbf{\Gamma}^e\|_2 \le \max_{1\le i\le p}\sum_{m=1}^p |\hat{\mathbf{\Gamma}}^{e\mathcal{T}}_{im}(r)-\mathbf{\Gamma}^e_{im}|$$
$$= \max_{1\le i\le p}\sum_{m=1}^p |\hat{\mathbf{\Gamma}}^{e\mathcal{T}}_{im}(r)-\mathbf{\Gamma}^e_{im}|\Big(I(|\mathbf{\Gamma}^e_{im}|\le\lambda_{pn})+I(|\mathbf{\Gamma}^e_{im}|>\lambda_{pn},|\hat{\mathbf{\Gamma}}^{e\mathcal{T}}_{im}(r)|<\lambda_{pn})$$
$$+I(|\mathbf{\Gamma}^e_{im}|>\lambda_{pn},|\hat{\mathbf{\Gamma}}^{e\mathcal{T}}_{im}(r)|>\lambda_{pn})\Big)$$
$$\le \max_{1\le i\le p}((C+1)+(C+1)2+((C+2)^q+1))\,\lambda_{pn}^{1-q}\sum_{m=1}^p|\mathbf{\Gamma}^e_{im}|^q \le C_q s_0(p)\lambda_{pn}^{1-q}. \quad (6.26)$$

(4.5) is now a direct consequence of (6.26) and (4.3). The proof of (4.6) is almost the same and hence omitted.

Proof of Theorem 4 Without confusion and for simplicity, we write $\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}}(\hat{r})$ in short by $\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}}$. Without loss of generality and for notational simplicity, we assume that $s=0$. The proof for $s\neq 0$ is almost the same. Recall that

$$\mathbf{\Sigma}_s = \boldsymbol{\sigma}_s\boldsymbol{\sigma}_s' + \boldsymbol{\sigma}_s^*\boldsymbol{\rho}^*\boldsymbol{\sigma}_s^{*\prime},\ \hat{\mathbf{\Sigma}}_s = \hat{\boldsymbol{\sigma}}_s\hat{\boldsymbol{\sigma}}_s' + \hat{\mathbf{\Sigma}}_s^{e\mathcal{T}}. \tag{6.27}$$

Let $\tilde{\mathbf{\Sigma}}_s^c = \tilde{\boldsymbol{\sigma}}_s\tilde{\boldsymbol{\sigma}}_s'$ where $\tilde{\boldsymbol{\sigma}}_s = \boldsymbol{\sigma}_s\boldsymbol{H}_s^{-1}$, and $\tilde{\mathbf{\Sigma}}_s = \tilde{\mathbf{\Sigma}}_s^c + \mathbf{\Sigma}_s^e$. Then we have

$$\|\hat{\mathbf{\Sigma}}_s^{-1} - \mathbf{\Sigma}_s^{-1}\| \le \|\hat{\mathbf{\Sigma}}_s^{-1} - \tilde{\mathbf{\Sigma}}_s^{-1}\| + \|\tilde{\mathbf{\Sigma}}_s^{-1} - \mathbf{\Sigma}_s^{-1}\|. \tag{6.28}$$

Now by the Sherman-Morrison-Woodbury formula, we have

$$\begin{aligned}
& \hat{\mathbf{\Sigma}}_s^{-1} - \tilde{\mathbf{\Sigma}}_s^{-1} \\
= \ & (\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1} - (\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s\left(I_r + \hat{\boldsymbol{\sigma}}_s'(\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s\right)^{-1}\hat{\boldsymbol{\sigma}}_s'(\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1} \\
& -\{(\mathbf{\Sigma}_s^e)^{-1} - (\mathbf{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s\left(I_r + \tilde{\boldsymbol{\sigma}}_s'(\mathbf{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s\right)^{-1}\tilde{\boldsymbol{\sigma}}_s'(\mathbf{\Sigma}_s^e)^{-1}\} \\
= \ & (\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1} - (\mathbf{\Sigma}_s^e)^{-1} - (\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s\left(I_r + \hat{\boldsymbol{\sigma}}_s'(\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s\right)^{-1}\hat{\boldsymbol{\sigma}}_s'(\hat{\mathbf{\Sigma}}_s^{e\mathcal{T}})^{-1} \\
& +(\mathbf{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s\left(I_r + \tilde{\boldsymbol{\sigma}}_s'(\mathbf{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s\right)^{-1}\tilde{\boldsymbol{\sigma}}_s'(\mathbf{\Sigma}_s^e)^{-1} \equiv VIII + IX + XI.
\end{aligned} \tag{6.29}$$

By Corollary 1, we have $VIII = O_p(\tilde{\lambda}_{pn}^{1-q}s_0(p))$. By (6.5), we have

$$\begin{aligned}
& \|\hat{\boldsymbol{\sigma}}_s - \tilde{\boldsymbol{\sigma}}_s\|^2 \\
\le \ & \|VI_s + \frac{\boldsymbol{\delta}_s - \overline{\boldsymbol{\delta}}_s}{k_n}\hat{\boldsymbol{F}}_s'\|_F^2 + \frac{\|\boldsymbol{\sigma}_s^*\boldsymbol{F}_s^*\hat{\boldsymbol{F}}_s'\|^2}{k_n} \equiv \zeta_1 + \zeta_2,
\end{aligned} \tag{6.30}$$

where $\|\boldsymbol{A}\|_F = \sqrt{tr(\boldsymbol{A}'\boldsymbol{A})}$. For ζ_1, we have

$$\|VI_s + \frac{\boldsymbol{\delta}_s - \overline{\boldsymbol{\delta}}_s}{k_n}\hat{\boldsymbol{F}}_s'\|_F^2 \le \|\boldsymbol{\sigma}_s\|_F^2\|\frac{\boldsymbol{F}_s - \boldsymbol{H}_s^{-1}\hat{\boldsymbol{F}}_s}{\sqrt{k_n}}\|_F^2\|\frac{\hat{\boldsymbol{F}}_s'}{\sqrt{k_n}}\|_F^2 + \|\frac{\boldsymbol{\delta}_s - \overline{\boldsymbol{\delta}}_s}{\sqrt{k_n}}\|_F^2\|\frac{\hat{\boldsymbol{F}}_s'}{\sqrt{k_n}}\|_F^2.$$

Let $\lambda_{pn}' = \frac{1}{p} + k_n\Delta_n$. By Lemma 2, the Burkhölder-Davis-Gundy inequality, the boundedness of μ_i's, and the fact that $\|\frac{\hat{\boldsymbol{F}}_s'}{\sqrt{k_n}}\|_F^2 \le r$, we have $\zeta_1 = O_p(\lambda_{pn}'p)$. For ζ_2, we have

$$\zeta_2 = \frac{1}{k_n^2}\lambda_{max}\left(\boldsymbol{\sigma}_s^*\boldsymbol{F}_s^*\hat{\boldsymbol{F}}_s'\hat{\boldsymbol{F}}_s\boldsymbol{F}_s^{*\prime}\boldsymbol{\sigma}_s^{*\prime}\right) = \|\boldsymbol{\sigma}_s^*\boldsymbol{F}_s^*\boldsymbol{F}_s^{*\prime}\boldsymbol{\sigma}_s^{*\prime}\|^{1/2}\frac{1}{k_n} \tag{6.31}$$

Then by (6.47) below, we have $\zeta_2 = O_p(\frac{p^{1/4}}{\sqrt{k_n}})$. So, in combination, we have

$$\|\hat{\boldsymbol{\sigma}}_s - \tilde{\boldsymbol{\sigma}}_s\|^2 = O_p(\lambda_{pn}'p). \tag{6.32}$$

Next, we have by the condition given in the theorem,

$$\begin{aligned}\lambda_{min}(G) &\equiv \lambda_{min}(I_r+\tilde{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s)\geq \lambda_{min}((\boldsymbol{\Sigma}_s^e)^{-1})\lambda_{min}(\tilde{\boldsymbol{\sigma}}_s'\tilde{\boldsymbol{\sigma}}_s)\\ &\geq C\lambda_{min}((\boldsymbol{H}_s^{-1})'\boldsymbol{\sigma}_s'\boldsymbol{\sigma}_s\boldsymbol{H}_s^{-1})\geq Cp, \end{aligned} \tag{6.33}$$

$$\text{and } \|\tilde{\boldsymbol{\sigma}}_s\|\leq Cp^{1/2}, \tag{6.34}$$

with probability approaching one by Corollary 2 and Assumptions 4-5. Let $G_1 = I_r+\hat{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s$, then by (6.32) and Corollary 1, we also have $\lambda_{min}(G_1)\geq Cp$ with probability approaching one. On the other hand, by (6.32) (6.33) and (6.34),

$$\begin{aligned}\|G^{-1}-G_1^{-1}\| &\leq \|G^{-1}\|\|G_1^{-1}\|\|\tilde{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s-\hat{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s\|\\ &= O_p(\frac{(\lambda_{pn}')^{1/2}}{p}+\frac{1}{k_n^{1/4}p^{11/8}}+\frac{\tilde{\lambda}_{pn}^{1-q}s_0(p)}{p}).\end{aligned} \tag{6.35}$$

Now we have the decomposition $IX+XI=\sum_{i=1}^5 L_i$, where

$$\begin{aligned}L_1 &= [(\boldsymbol{\Sigma}_s^e)^{-1}-(\hat{\boldsymbol{\Sigma}}_s^{e\mathcal{T}})^{-1}]\tilde{\boldsymbol{\sigma}}_s G^{-1}\tilde{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\\ L_2 &= (\hat{\boldsymbol{\Sigma}}_s^{e\mathcal{T}})^{-1}[\tilde{\boldsymbol{\sigma}}_s-\hat{\boldsymbol{\sigma}}_s]G^{-1}\tilde{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\\ L_3 &= (\hat{\boldsymbol{\Sigma}}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s G^{-1}[\tilde{\boldsymbol{\sigma}}_s'-\hat{\boldsymbol{\sigma}}_s](\boldsymbol{\Sigma}_s^e)^{-1}\\ L_4 &= (\hat{\boldsymbol{\Sigma}}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s G^{-1}\hat{\boldsymbol{\sigma}}_s[(\boldsymbol{\Sigma}_s^e)^{-1}-(\hat{\boldsymbol{\Sigma}}_s^{e\mathcal{T}})^{-1}]\\ L_5 &= (\hat{\boldsymbol{\Sigma}}_s^{e\mathcal{T}})^{-1}\hat{\boldsymbol{\sigma}}_s[G^{-1}-G_1^{-1}]\hat{\boldsymbol{\sigma}}_s(\hat{\boldsymbol{\Sigma}}_s^{e\mathcal{T}})^{-1}\end{aligned}$$

By this decomposition and (6.32) (6.33) (6.34) and (6.35), we have $IX+XI= O_p(\tilde{\lambda}_{pn}^{1-q}s_0(p)+\sqrt{\lambda_{pn}'}+\frac{1}{k_n^{1/4}p^{3/8}})$. This together with the estimate of $VIII$, we show that $\hat{\boldsymbol{\Sigma}}_s^{-1}-\tilde{\boldsymbol{\Sigma}}_s^{-1}=O_p(\lambda_{pn}^{1-q}s_0(p)+\sqrt{\lambda_{pn}'}+\frac{1}{k_n^{1/4}p^{3/8}})$. Similarly, by the Sherman-Morrison-Woodbury formula again, we have

$$\begin{aligned}\tilde{\boldsymbol{\Sigma}}_s^{-1}-\boldsymbol{\Sigma}_s^{-1} &= (\boldsymbol{\Sigma}_s^e)^{-1}\boldsymbol{\sigma}_s\left(I_r+\boldsymbol{\sigma}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\boldsymbol{\sigma}_s\right)^{-1}\boldsymbol{\sigma}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\\ &\quad -(\boldsymbol{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s(I_r+\tilde{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s)^{-1}\tilde{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^e)^{-1}.\end{aligned}$$

Let $\tilde{G}=I_r+\tilde{\boldsymbol{\sigma}}_s'(\boldsymbol{\Sigma}_s^e)^{-1}\tilde{\boldsymbol{\sigma}}_s$ and $\tilde{G}_1=I_r+\boldsymbol{\sigma}_s'(\boldsymbol{\Sigma}^e)_s^{-1}\boldsymbol{\sigma}_s$. Then we have

$$\lambda_{min}(\tilde{G})\geq \lambda_{min}((\boldsymbol{\Sigma}_s^e)^{-1})\lambda_{min}(\tilde{\boldsymbol{\sigma}}_s'\tilde{\boldsymbol{\sigma}}_s)\geq Cp,$$

with probability approaching one due to Assumption 4, the condition in the theorem, and Corollary 2. Similarly, $\lambda_{min}(\tilde{G}_1)\geq Cp$ with probability approaching one. Therefore, by (6.11),

$$\|\tilde{G}_1^{-1}-\boldsymbol{H}_s^{-1}\tilde{G}^{-1}\boldsymbol{H}_s^{-1}\|\leq C\|\tilde{G}^{-1}\|\|\tilde{G}_1^{-1}\|\|\boldsymbol{H}_s\boldsymbol{H}_s-I_r\|=O_p(\frac{1}{p^{9/4}}+\frac{1}{p^2}\sqrt{\frac{\log p}{k_n^{1-\epsilon}}}).$$

This together with the fact that $\|\boldsymbol{\sigma}_s\| \le C\sqrt{p}$ yields $\tilde{\boldsymbol{\Sigma}}_s^{-1} - \boldsymbol{\Sigma}_s^{-1} = O_p(\frac{1}{p^{5/4}} + \frac{1}{p}\sqrt{\frac{\log p}{k_n^{1-\epsilon}}})$. This completes the proof.

Appendix B: Proof of the Lemmas

Proof of Lemma 1 *proof of (6.9)* By Weyl's theorem, we have

$$\max_{1\le j\le r} |v_j - v_j^0| \le C(\|\frac{\boldsymbol{\delta}_k'\boldsymbol{\delta}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k}{pk_n}\| + \|\frac{\boldsymbol{\sigma}_k(\boldsymbol{F}_k\boldsymbol{F}_k'/k_n)\boldsymbol{\sigma}_k' - \boldsymbol{\sigma}_k\boldsymbol{\sigma}_k'}{p}\|). \tag{6.36}$$

By Assumption 5, we only need to prove

$$P\left(\|\frac{\boldsymbol{\delta}_k'\boldsymbol{\delta}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k}{pk_n}\| > h(\frac{1}{p^{1/4}} + \frac{(\log p)^{1/4}}{k_n^{1/2-\epsilon}})\right) = O(p^{-\frac{\delta'}{2}\vee(-M)}), \tag{6.37}$$

$$\text{and } P\left(\|\boldsymbol{F}_k\boldsymbol{F}_k'/k_n - I_r\| > h\sqrt{\frac{\log p}{k_n}}\right) = O(p^{-M}), \tag{6.38}$$

for any $M > 0$ if h is chosen large. For (6.38), notice that the difference is a fixed dimensional matrix, it suffices to prove

$$P\left(|\sum_{j=1}^{k_n} F(i,j)F(l,j)/k_n - I_{il}| > h\sqrt{\frac{\log p}{k_n}}\right) = O(p^{-M}), \tag{6.39}$$

where $I_{il} = 1$ if $i = l$ and 0 otherwise. By the Markov inequality and the Gaussianity of $F(i,j)$'s,

$$P\left(\frac{\sum_{j=1}^{k_n} F(i,j)F(l,j)}{k_n} - I_{il} > h\sqrt{\frac{\log p}{k_n}}\right) \le e^{-xh\sqrt{\frac{\log p}{k_n}} + Cx^2/k_n}.$$

Similarly, we have

$$P\left(\frac{\sum_{j=1}^{k_n} F(i,j)F(l,j)}{k_n} - I_{il} < -h\sqrt{\frac{\log p}{k_n}}\right) \le e^{-xh\sqrt{\frac{\log p}{k_n}} + Cx^2/k_n}.$$

Taking $x = \sqrt{k_n \log p}$ in last two equations proves (6.39). For (6.37), we bound the quantity within the parentheses in the left hand side by

$$\frac{\|(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)'(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)\| + 2\|\overline{\boldsymbol{\delta}}_k'(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)\| + \|\overline{\boldsymbol{\delta}}_k'\overline{\boldsymbol{\delta}}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k\|}{pk_n}. \tag{6.40}$$

We further have

$$\frac{\|(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)'(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)\|}{pk_n} \le \frac{1}{k_n}\sum_{j=1}^{k_n}\frac{1}{p}\sum_{i=1}^{p}(\delta_k(i,j) - \overline{\delta}_k(i,j))^2$$
$$\le \max_{1\le i\le p}\frac{1}{k_n}\sum_{j=1}^{k_n}(\delta_k(i,j) - \overline{\delta}_k(i,j))^2.$$

By the above equation, boundedness of μ_i's and (S.3.49) in the supplement, we have for any $\epsilon > 0$ and large h,

$$P\left(\frac{\|(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)'(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)\|}{pk_n} > h\frac{\sqrt{\log p}}{k_n^{1-\epsilon}}\right) = O(p^{-M}). \tag{6.41}$$

By the above equation, Assumptions 3 and 5, (6.38) and

$$\frac{\|\overline{\boldsymbol{\delta}}_k'(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)\|}{pk_n} \le \frac{\|\overline{\boldsymbol{\delta}}_k\|}{\sqrt{pk_n}}\frac{\|\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k\|}{\sqrt{pk_n}}$$

we have

$$P\left(\frac{\|\overline{\boldsymbol{\delta}}_k'(\boldsymbol{\delta}_k - \overline{\boldsymbol{\delta}}_k)\|}{pk_n} > h\frac{(\log p)^{1/4}}{k_n^{1/2-\epsilon}}\right) = O(p^{-M}). \tag{6.42}$$

$$\|\overline{\boldsymbol{\delta}}_k'\overline{\boldsymbol{\delta}}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k\| \le \|\boldsymbol{F}_k^{*\prime}\boldsymbol{\sigma}_k^{*\prime}\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\| + 2\|\boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\|. \tag{6.43}$$

Since $\boldsymbol{\sigma}_k^*$ is a diagonal matrix with bounded entries as assumed in Assumption 5,

$$\|\boldsymbol{F}_k^{*\prime}\boldsymbol{\sigma}_k^{*\prime}\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\| \le C\|\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^*\| \le C\left(\|\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^* - E[\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^*]\| + \|E[\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^*]\|\right). \tag{6.44}$$

Since $E[\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^*]$ is a diagonal matrix, we have $\|E[\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^*]\| \le C$ and hence

$$\|\boldsymbol{F}_k^{*\prime}\boldsymbol{\sigma}_k^{*\prime}\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\|/pk_n \le C(1 + \|\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^* - E[\boldsymbol{F}_k^{*\prime}\boldsymbol{F}_k^*]\|)/pk_n$$
$$\le C(1/pk_n + \sqrt{\frac{1}{pk_n^2}\sum_{j=1}^{k_n}\sum_{l=1}^{k_n}\left(\frac{\boldsymbol{F}_k^{*\prime}(j)\boldsymbol{F}_k^*(l) - E[\boldsymbol{F}_k^{*\prime}(j)\boldsymbol{F}_k^*(l)]}{\sqrt{p}}\right)^2}). \tag{6.45}$$

Let $\boldsymbol{F}_k^* = \boldsymbol{\rho}^{*\frac{1}{2}}\tilde{\boldsymbol{F}}_k^*$ where $\tilde{\boldsymbol{F}}_k^*$ has independent entries. Then we have

$$E\left[\frac{1}{k_n^2}\sum_{j=1}^{k_n}\sum_{l=1}^{k_n}\left(\frac{\boldsymbol{F}_k^{*\prime}(j)\boldsymbol{F}_k^*(l) - E[\boldsymbol{F}_k^{*\prime}(j)\boldsymbol{F}_k^*(l)]}{\sqrt{p}}\right)^2\right]^{\delta'} \tag{6.46}$$
$$\le \max_{j,l} E\left(\frac{\boldsymbol{F}_k^{*\prime}(j)\boldsymbol{F}_k^*(l) - E[\boldsymbol{F}_k^{*\prime}(j)\boldsymbol{F}_k^*(l)]}{\sqrt{p}}\right)^{2\delta'} \le C\left(\frac{\sum_i\sum_j\rho_{ij}^{*2}}{p}\right)^{\delta'}.$$

By this and Assumption 3, plus (6.45), we have with the Markov inequality,

$$P\left(\frac{\|\boldsymbol{F}_k^{*\prime}\boldsymbol{\sigma}_k^{*\prime}\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\|}{pk_n} > \frac{h}{p^{1/4}}\right) = O(p^{-\delta'/2}). \tag{6.47}$$

Simple manipulation yields,

$$\frac{\|\boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\|}{pk_n} \leq \sqrt{\frac{1}{pk_n^2}\sum_{m=1}^{k_n}\sum_{j=1}^{k_n}\left[\sum_{i=1}^{p}\frac{F_k^*(i,j)}{\sqrt{p}}\sigma_{ki}^*\sum_{l=1}^{r}\sigma_k(i,l)F_k(l,m)\right]^2}. \tag{6.48}$$

Notice that $\sum_{i=1}^{p}\frac{F_k^*(i,j)}{\sqrt{p}}\sigma_{ki}^*\sum_{l=1}^{r}\sigma_k(i,l)F_k(l,m)$ is a Gaussian random variable conditional on $\mathcal{F}_{k-1}$ and $\boldsymbol{F}_k$,

$$E\left[\sum_{i=1}^{p}\frac{F_k^*(i,j)}{\sqrt{p}}\sigma_{ki}^*\sum_{l=1}^{r}\sigma_k(i,l)F_k(l,m)\right]^{2\delta'} \leq C\left(\frac{\sum_i\sum_j|\rho_{ij}^*|}{p}\right)^{2\delta'}. \tag{6.49}$$

By again the Markov inequality and Assumption 3, we have

$$P\left(\frac{\|\boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k^*\boldsymbol{F}_k^*\|}{pk_n} > \frac{h}{p^{1/4}}\right) = O(p^{-\delta'/2}). \tag{6.50}$$

(6.47) and (6.50) prove that

$$P\left(\frac{\|\overline{\boldsymbol{\delta}}_k'\overline{\boldsymbol{\delta}}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k\|}{pk_n} > \frac{h}{p^{1/4}}\right) = O(p^{-\delta'/2}). \tag{6.51}$$

(6.41) (6.42) and (6.51) proves (6.37).

proof of (6.10) We first give a representation of $\frac{\boldsymbol{F}_k\hat{\boldsymbol{F}}_k'}{k_n}$. Let $\boldsymbol{R}_k = (\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}\frac{\boldsymbol{F}_k\hat{\boldsymbol{F}}_k'}{k_n}$, $\boldsymbol{d}_k = \frac{1}{k_n}(\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}\boldsymbol{F}_k\frac{\boldsymbol{\delta}_k'\boldsymbol{\delta}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k}{pk_n}\hat{\boldsymbol{F}}_k'$, $\boldsymbol{B}_k = (\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}\frac{\boldsymbol{F}_k\boldsymbol{F}_k'}{k_n}(\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}$, $\boldsymbol{L}_{nk}^* = (diag(\boldsymbol{R}_k'\boldsymbol{R}_k))^{-1/2}$, and $\boldsymbol{\gamma}_{nk} = \boldsymbol{R}_k(diag(\boldsymbol{R}_k'\boldsymbol{R}_k))^{-1/2}$. From the equality that $\frac{\boldsymbol{\delta}_k'\boldsymbol{\delta}_k}{pk_n}\hat{\boldsymbol{F}}_k' = \hat{\boldsymbol{F}}_k'\boldsymbol{V}_k$, we deduce that

$$\left(\boldsymbol{B}_k + \boldsymbol{d}_k\boldsymbol{R}_k^{-1}\right)\boldsymbol{\gamma}_{nk} = \boldsymbol{\gamma}_{nk}\boldsymbol{V}_k, \tag{6.52}$$

demonstrating that $\boldsymbol{\gamma}_{nk}$ is the eigenvector matrix of $\boldsymbol{B}_k + \boldsymbol{d}_k\boldsymbol{R}_k^{-1}$. Then from the definition of $\boldsymbol{R}_k$, we have

$$\frac{\boldsymbol{F}_k\hat{\boldsymbol{F}}_k'}{k_n} = (\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{-1/2}\boldsymbol{\gamma}_{nk}\boldsymbol{L}_{nk}^*, \tag{6.53}$$

$$\text{and } \boldsymbol{H}_k = \boldsymbol{V}_k^{-1}\boldsymbol{L}_{nk}^*\boldsymbol{\gamma}_{nk}'(\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}. \tag{6.54}$$

Seen from (6.9) and (6.54), it is enough to show that

$$P\left(\|\boldsymbol{\gamma}_{nk}-\boldsymbol{\gamma}^0_k\|>h(\frac{\sqrt{\log p}}{\sqrt{k_n}}+\frac{1}{\sqrt{k_n}p^{1/4}})\right)=O(p^{-\frac{\delta'}{2}\vee(-M)}), \tag{6.55}$$

$$\text{and } P\left(\|\boldsymbol{L}^{*2}_{nk}-\boldsymbol{V}^0_k\|>h(\frac{\sqrt{\log p}}{\sqrt{k_n}}+\frac{1}{p^{1/4}})\right)=O(p^{-\frac{\delta'}{2}\vee(-M)}). \tag{6.56}$$

Let $\boldsymbol{\gamma}_k$ be the eigenvector matrix of $\boldsymbol{B}_k$. Then we have by the $SIN(\theta)$ Theorem (Davis and Kahan (1970)),

$$\|\boldsymbol{\gamma}_{nk}-\boldsymbol{\gamma}_k\|\leq C\|\boldsymbol{R}_k^{-1}\|\|\boldsymbol{d}_k\|. \tag{6.57}$$

By Assumption 4, the definition of $\hat{\boldsymbol{F}}_k$, and (6.38), we have, for large enough h' and C and some $C^*, c, c'>0$,

$$\begin{aligned} P\left(\|\boldsymbol{R}_k^{-1}\|>h'\right) &\leq P\left(\lambda_{min}(\frac{\boldsymbol{F}_k\boldsymbol{F}'_k}{k_n})\leq C^{-1}\right)\\ &\leq P\left(1-C^*\|\frac{\boldsymbol{F}_k\boldsymbol{F}'_k}{k_n}-I_r\|\leq C^{-1}\right)\\ &\leq P\left(\|\frac{\boldsymbol{F}_k\boldsymbol{F}'_k}{k_n}-I_r\|\geq\frac{c}{2}\right)=O(e^{-c'k_n})=O(p^{-M}), \end{aligned} \tag{6.58}$$

due to $p=e^{o(k_n)}$. On the other hand,

$$\|\boldsymbol{d}_k\|\leq\frac{1}{\sqrt{k_n}}\|(\frac{\boldsymbol{\sigma}'_k\boldsymbol{\sigma}_k}{p})^{1/2}\|\frac{\|\boldsymbol{\delta}'_k\boldsymbol{\delta}_k-\boldsymbol{F}'_k\boldsymbol{\sigma}'_k\boldsymbol{\sigma}_k\boldsymbol{F}_k\|}{pk_n} \tag{6.59}$$

By (6.37), we have

$$P\left(\|\boldsymbol{d}_k\|>h(\frac{\sqrt{\log p}}{k_n^{1-\epsilon}}+\frac{1}{p^{1/4}k_n^{1/2}})\right)=O(p^{-\frac{\delta'}{2}\vee(-M)}). \tag{6.60}$$

Combining (6.58) and (6.60) proves that

$$P\left(\|\boldsymbol{\gamma}_{nk}-\boldsymbol{\gamma}_k\|>h(\frac{\sqrt{\log p}}{k_n^{1-\epsilon}}+\frac{1}{p^{1/4}k_n^{1/2}})\right)=O(p^{-\frac{\delta'}{2}\vee(-M)}). \tag{6.61}$$

By the $SIN(\theta)$ Theorem again and (6.38), we have

$$P\left(\|\boldsymbol{\gamma}_k-\boldsymbol{\gamma}^0_k\|>h\sqrt{\frac{\log p}{k_n}}\right)=O(p^{-\frac{\delta'}{2}\vee(-M)}). \tag{6.62}$$

(6.61) and (6.62) prove (6.55). Now we proceed to prove (6.56). Let $\boldsymbol{L}_k$ be the eigenvalue matrix of $\boldsymbol{B}_k$, notice that $\frac{\hat{\boldsymbol{F}}_k(\frac{\boldsymbol{\delta}_k'\boldsymbol{\delta}_k}{pk_n})\hat{\boldsymbol{F}}_k'}{k_n} = \boldsymbol{V}_k$, we have

$$\begin{aligned}
&\|\frac{\hat{\boldsymbol{F}}_k'\boldsymbol{F}_k'}{k_n}\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p}\frac{\boldsymbol{F}_k\hat{\boldsymbol{F}}_k'}{k_n} - \boldsymbol{V}_k^0\| \leq \|\frac{\hat{\boldsymbol{F}}_k'\boldsymbol{F}_k'}{k_n}\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p}\frac{\boldsymbol{F}_k\hat{\boldsymbol{F}}_k'}{k_n} - \boldsymbol{V}_k\| + \|\boldsymbol{V}_k - \boldsymbol{V}_k^0\| \\
\leq\ & \frac{C\|\boldsymbol{\delta}_k'\boldsymbol{\delta}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k\|}{pk_n} + \|\boldsymbol{V}_k - \boldsymbol{V}_k^0\|
\end{aligned}$$

(6.63)

This together with (6.9) and (6.37) proves (6.56).
proof of (6.11) (6.11) is a direct consequence of (6.10) because

$$I_r = (\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}\boldsymbol{\gamma}_k^0(\boldsymbol{V}_k^0)^{-1}\boldsymbol{\gamma}_k^{0\prime}(\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^{1/2}.$$

Proof of Corollary 2 Following the proof of Lemma 1 by simply changing $\frac{1}{p^{1/4}}$, and $\frac{\sqrt{\log p}}{k_n^{1/2-\epsilon}}$ by generic constants, we have

$$P\left(\max_k(\|\boldsymbol{V}_k\| + \|\boldsymbol{V}_k^{-1}\| + \|\boldsymbol{H}_k\| + \|\boldsymbol{H}_k^{-1}\|) > h'\right) = O(p^{-\delta'}k_n) = o(1).$$

By (6.4),

$$\frac{\|\hat{\boldsymbol{F}}_k - \boldsymbol{H}_k\boldsymbol{F}_k\|}{\sqrt{k_n}} \leq \|\boldsymbol{V}_k^{-1}\|\|\frac{\boldsymbol{\delta}_k'\boldsymbol{\delta}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k}{pk_n}\|.$$

Then by following the proof of (6.37) by changing $1/p^{1/4}$ by generic constants, we have

$$P\left(\frac{\|\hat{\boldsymbol{F}}_k - \boldsymbol{H}_k\boldsymbol{F}_k\|}{\sqrt{k_n}} > h'\right) = O(p^{-\delta'}k_n) = o(1).$$

Proof of Lemma 2 (6.13) is a straightforward result of the independence and Gaussianity of $\boldsymbol{F}_k(j)$'s and (6.58). By Assumption 5 and the Burkhölder-Davis-Gundy inequality, we have

$$E_{\mathcal{F}_{k-1}}\left(\frac{\|(\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}}_k)'(\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}}_k)\|}{pk_n}\right)^d \leq E_{\mathcal{F}_{k-1}}\left(\frac{tr(\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}}_k)'(\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}}_k)}{pk_n}\right)^d \leq C(k_n\Delta_n)^d,$$

$$\text{and } E_{\mathcal{F}_{k-1}}\left(\frac{\|\bar{\boldsymbol{\delta}}_k'(\boldsymbol{\delta}_k - \bar{\boldsymbol{\delta}}_k)\|}{pk_n}\right)^d \leq C(k_n\Delta_n)^{d/2}.$$

By (6.43)-(6.46) and (6.48)-(6.49), we have

$$E_{\mathcal{F}_{k-1}}\left(\frac{\bar{\boldsymbol{\delta}}_k\bar{\boldsymbol{\delta}}_k' - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k}{pk_n}\right)^d \leq Cp^{-d/2}.$$

Combining these facts, by (6.40), the bound of $\|\boldsymbol{\delta}_k'\boldsymbol{\delta}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k\|/(pk_n)$, we have $E_{\mathcal{F}_{k-1}}(\|\boldsymbol{\delta}_k'\boldsymbol{\delta}_k - \boldsymbol{F}_k'\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k\boldsymbol{F}_k\|/(pk_n))^d \leq C(\frac{1}{C_{pn}})^d$. Due to the Gaussianity of $\boldsymbol{F}_k$ and Assumption 4, $E_{\mathcal{F}_{k-1}}(\|\boldsymbol{\sigma}_k(\boldsymbol{F}_k\boldsymbol{F}_k'/k_n - I_r)\boldsymbol{\sigma}_k'\|/p)^d \leq Ck_n^{-d/2}$ and hence by (6.36), $E_{\mathcal{F}_{k-1}}(\|\boldsymbol{V}_k - \boldsymbol{V}_k^0\| + \|\boldsymbol{V}_k^{-1} - (\boldsymbol{V}_k^0)^{-1}\|)^d \leq C(C_{pn})^d$. This proves part of (6.15). Then by Assumption 4, $E_{\mathcal{F}_{k-1}}(\|\boldsymbol{V}_k\| + \|\boldsymbol{V}_k^{-1}\|)^d \leq C$. This proves part of (6.15). Following the proof of (6.10), we have $E_{\mathcal{F}_{k-1}}(\|\boldsymbol{H}_k - \boldsymbol{H}_k^0\|^d + \|\boldsymbol{H}_k^{-1} - (\boldsymbol{H}_k^0)^{-1}\|^d) \leq C(\frac{1}{C_{pn}})^d$ where $\boldsymbol{H}_k^0 = (\boldsymbol{V}_k^0)^{-1/2}\gamma_k^0(\frac{\boldsymbol{\sigma}_k'\boldsymbol{\sigma}_k}{p})^2$. This proves the remaining part of (6.15). This implies $E_{\mathcal{F}_{k-1}}(\|\boldsymbol{H}_k\| + \|\boldsymbol{H}_k^{-1}\|)^d \leq C$. This proves the remaining part of (6.15). (6.16) is simply (6.31) of Lemma 2 in Kong (2016a).

ACKNOWLEDGEMENTS

We want to thank the Associate Editor and two referees for their constructive comments and remarks that lead to the improvement of the paper.

SUPPLEMENTARY MATERIAL

Supplement: Supplement to "On the integrated systematic and idiosyncratic volatility with large panel high-frequency data"
(DOI:). This supplement contains the technical proof of Lemmas 3-5, which is crucial in proving Theorem 1 and Theorem 2.

REFERENCES

Aït-Sahalia, Y. and Xiu, D. (2016a). Principal component analysis of high frequency data. Manuscript.

Aït-Sahalia, Y. and Xiu, D. (2016b). Using principal component analysis to estimate a high dimensional factor model with high-frequency data, *Journal of Econometrics* to appear.

Andersen, T. G., Bollerslev, T., Diebold, F. and Labys, P. (2003). Modeling and Forecasting Realized Volatility, *Econometrica 71*, 579-625.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135-171.

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.

Barndorff-Nielsen, O. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of Royal Statistical Society, Series B, 64*, 253-280.

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

Cai, T., and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association, 106*, 672-684.

Chamberlain, G., and Rothschild, M. (1983). Arbitarge, factor structure, and mean-variance analysis on large asset markets. *Econometrica, 51*, 1281-1304.

Chen, S. X. and Xu, Z. (2014). On Implied Volatility for Options - Some Reasons to Smile and More to Correct. *Journal of Econometrics, 179*, 1-15.

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010), Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics 159*, 116-133.

Davis and Kahan (1970). The rotation of eigenvectors by a perturbation, III^*. *SIAM Journal of Numerical Analysis 7*, 1-46.

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

Fan, J., Li, Y. and Ke, Y. (2012). Vast volatility matrix estimation using high frequency data for portfolio selection. *Journal of American Statistical Association 107*, 412-428.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B, 75* 603-680.

Jacod, J., and Rosenbaum, M. (2013). Quarticity and other functionals of volatility: efficient estimation. *Annals of Statistics 41*, 1462-1484.

Jacod, J., and Todorov, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Annals of Statistics 42*, 1029-1069.

Jing, B. Y., Kong, X. B., Liu, Z. and Mykland, P. (2012a). On the jump activity index on semimartingales. *Journal of Econometrics 166*, 213-223.

Jing, B. Y., Kong, X. B. and Liu, Z. (2012b). Modeling high frequency financial data by pure jump models. *Annals of Statistics 40*, 759-784.

Kim, D., Kong, X. B., Li, C., and Wang, Y. (2015). Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data. Under Review.

Kong, X. B. (2016a). Determining the number of common driving Brownian motions underlying high frequency data? http://ssrn.com/abstract=2752456.

Kong, X. B. (2016b). Supplement to "On the integrated systematic and idiosyncratic volatility with large panel high-frequency data". DOI:

Kong, X. B., Liu, Z. and Jing, B. Y. (2015). Testing for pure-jump processes for high-frequency data. *Annals of Statistics 43*, 847-877.

Li, J., and Chen, S. X. (2012). Two sample tests for high dimensional covariance matrices. *Annals of Statistics, 40*, 908-940.

Mykland, P. A., Shephard, N. and Shephard, K. (2012). Efficient and feasible inference for the components of financial variation using blocked multipower variation. Technical report.

Mykland, P. A., and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequncy. *Econometrica 77*, 1403-1445.

Pelger, M. (2016). Large-dimensional factor modeling based on high-frequency observations, Available at SSRN 2584172.

Qiu, Y. M. and Chen, S. X. (2012). Test for Bandedness of High Dimensional Covariance Matrices with Bandwidth Estimation, *Annals of Statistics, 40*, 1285-1314.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory, 13*, 341-360.

Tao, M., Wang, Y. and Chen, X. (2013a). Fast convergence rates in estimating large volatility matrices using high-frequency financial data. *Econometric Theory, 29*, 838-856.

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

Todorov, V. and Tauchen, G. (2012a). The realized Laplace transform of volatility. *Econometrica 80*, 1105-1127.

Todorov, V. and Tauchen, G. (2012b). Inverse Realized Laplace Transform for Nonparametric Volatility Density Estimation in Jump Diffusions, *Journal of the American Statistical Association 107*, 622-635.

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

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.

West Yushan Road 86,
Pukou, Nanjing, P. R. China
E-mail: kongxblqh@gmail.com