Published in final edited form as:

J Am Stat Assoc. 2018 ; 113(523): 1268–1283. doi:10.1080/01621459.2017.1340888.

Robust High-dimensional Volatility Matrix Estimation for High-Frequency Factor Model*

Jianqing Fan and Donggyu Kim

Princeton University

Abstract

High-frequency financial data allow us to estimate large volatility matrices with relatively short time horizon. Many novel statistical methods have been introduced to address large volatility matrix estimation problems from a high-dimensional Itô process with microstructural noise contamination. Their asymptotic theories require sub-Gaussian or some finite high-order moments assumptions for observed log-returns. These assumptions are at odd with the heavy tail phenomenon that is pandemic in financial stock returns and new procedures are needed to mitigate the influence of heavy tails. In this paper, we introduce the Huber loss function with a diverging threshold to develop a robust realized volatility estimation. We show that it has the sub-Gaussian concentration around the volatility with only finite fourth moments of observed log-returns. With the proposed robust estimator as input, we further regularize it by using the principal orthogonal component thresholding (POET) procedure to estimate the large volatility matrix that admits an approximate factor structure. We establish the asymptotic theories for such low-rank plus sparse matrices. The simulation study is conducted to check the finite sample performance of the proposed estimation methods.

Key words and phrases

Concentration inequality; Huber loss; low-rank matrix; pre-averaging; sparsity

1 Introduction

Ever since the wide availability of high-frequency financial data, researchers have devoted to developing well-performing procedures for estimating integrated volatilities. These include two-time scale realized volatility (TSRV) (Zhang et al., 2005), multi-scale realized volatility (MSRV) (Zhang, 2006, 2011), wavelet estimator (Fan and Wang, 2007), kernel realized volatility (KRV) (Barndorff-Nielsen et al., 2008, 2011), pre-averaging realized volatility (PRV) (Christensen et al., 2010; Jacod et al., 2009), a quasi-maximum likelihood estimator (QMLE) (Aït-Sahalia et al., 2010; Xiu, 2010), and local method of moments (Bibinger et al., 2014). When the number of assets is small, these estimation methods perform well and obtain the optimal convergence rate in presence of microstructural noises. However, in many financial studies, we encounter a large number of assets, and it is well-known that the

*Supported by NSF Grant DMS-1406266 and NIH Grant R01-GM072611-11, and National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100190.

estimation methods designed for finite-dimensional matrices are neither effective nor efficient. Recent studies of high-dimensional statistics indicate that some structures such as sparsity or factor modeling are required in order to estimate large covariance or volatility matrices effectively and efficiently (Fan et al., 2013). For example, Tao et al. (2013b) established the minimax lower bound for estimating sparse integrated volatility matrices and showed that the thresholding volatility estimator can achieve this rate. See also Kim et al. (2015a); Wang and Zou (2010). However, financial volatility matrices can not be sparse due to the existence of common factors and large correlation among assets, and so to better accommodate the sparsity, several estimation methods for factor-based high-dimensional Itô processes have been proposed (Ait-Sahalia and Xiu, 2015; Fan et al., 2016a; Kim et al., 2015b). These estimation methods empirically show better performance than the thresholding methods. In this paper, we use the factor-based Itô process to investigate the large volatility matrices.

To study the large volatility matrices, some sub-Gaussian or the finite high-order moment conditions have been imposed. The current large volatility matrix estimation methods have trade-offs between the convergence rate and the orders of finite moments. Unfortunately, these assumptions contradict with many well-documented stylized features of financial returns, which in fact exhibit heavy tails. Thus, the light-tail conditions on the asset returns can not be satisfied. This fact demands for developing a robust estimation method which can obtain the sub-Gaussian concentration with only finite fourth moments. For i.i.d. observations, Fan et al. (2016c) introduced the robust estimators by minimizing the Huber loss function (Huber, 1964) with a diverging threshold to reduce biases. This new estimator reduces the requirements from sub-Gaussian tails to only finite fourth moments. See also Catoni (2012); Minsker (2016); Mykland and Zhang (2016). However, these estimation methods and asymptotic theories are not directly applicable to the high-frequency financial data because observations are neither identical nor independent and the parameter of interest, the integrated volatility matrix, is a random variable. In addition, stock prices are contaminated by the microstructural noises, and can not be observed directly.

In this paper, we develop a robust realized volatility estimation procedure based on high-frequency financial data coming from an Itô process with noise contamination. This large robust volatility matrix is further regularized by using the principal orthogonal component thresholding (POET) procedure (Fan et al., 2013), which takes advantage of the low-rank plus sparse structure of the matrix. Specifically, since the observations are contaminated by the microstructural noises, we first apply a pre-averaging scheme to remove the effect of the microstructural noises. We then employ the Huber loss function with a diverging parameter in order to deal with heavy-tailed observations, where the heavy tail is coming from the microstructural noise. We use the minimizer of the Huber loss function as an estimator of the integrated volatility. We call the proposed robust estimator the robust pre-averaging realized volatility (RPRV) estimator. We show that the RPRV estimator has the sub-Gaussian concentration around the volatility with only finite fourth moments of observed log-returns. Such a concentration is fundamental in studying high-dimensional matrix inferences. This RPRV has not taken into account the structure of the volatility matrix. To estimate large volatility matrices, we introduce the factor-based Itô processes and assume that the integrated volatility matrix consists of low-rank and sparse matrices. The principal

orthogonal component thresholding (POET) procedure is employed to further regularize the RPRV. The asymptotic properties of our final estimator, the regularized RPRV, are also investigated.

2 The model set-up

Denote by $\mathbf{X}(t) = (X_1(t), \dots, X_p(t))^T$ the vector of true log-prices of p assets at time t . Common approach in modern finance theory assumes that the log-prices of assets follow a continuous-time diffusion model. To account for the effect of common factors, we use the following factor-based diffusion model,

$$d\mathbf{X}(t) = \boldsymbol{\mu}(t)dt + \boldsymbol{\vartheta}^T(t)d\mathbf{W}_t^* + \boldsymbol{\sigma}^T(t)d\mathbf{W}_t, \quad (2.1)$$

where $\boldsymbol{\mu}(t) = (\mu_1(t), \dots, \mu_p(t))^T$ is a drift vector, $\boldsymbol{\vartheta}(t)$ is a r by p factor loading matrix, $\boldsymbol{\sigma}(t)$ is a p by p matrix, stochastic processes $\boldsymbol{\mu}(t)$, $\mathbf{X}(t)$, $\boldsymbol{\sigma}(t)$, and $\boldsymbol{\vartheta}(t)$ are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in [0, 1]\}, P)$ with filtration $\mathcal{F}_t$ satisfying the usual conditions, $\mathbf{W}_t^*$ and $\mathbf{W}_t$ are r -dimensional and p -dimensional independent Brownian motions, respectively, and they are independent. The instantaneous (or spot) volatility matrix of the log-prices $\mathbf{X}(t)$ is then give by

$$\boldsymbol{\gamma}(t) = (\gamma_{ij}(t))_{1 \leq i, j \leq p} = \boldsymbol{\vartheta}^T(t)\boldsymbol{\vartheta}(t) + \boldsymbol{\sigma}^T(t)\boldsymbol{\sigma}(t),$$

and their quadratic variation is

$$[\mathbf{X}, \mathbf{X}]_t = \int_0^t \boldsymbol{\gamma}(s)ds = \left(\int_0^t \gamma_{ij}(s)ds \right)_{1 \leq i, j \leq p} = \int_0^t \boldsymbol{\vartheta}^T(s)\boldsymbol{\vartheta}(s)ds + \int_0^t \boldsymbol{\sigma}^T(s)\boldsymbol{\sigma}(s)ds.$$

The parameters of interest are the integrated volatility matrix

$$\boldsymbol{\Gamma} = [\mathbf{X}, \mathbf{X}]_1 = \int_0^1 \boldsymbol{\gamma}(s)ds = \boldsymbol{\Theta} + \boldsymbol{\Sigma}. \quad (2.2)$$

The matrix $\boldsymbol{\Theta}$ in (2.2) accounts for the influence of factors on the volatility matrix and its rank is typically small. It admits the following spectral-decomposition

$$\boldsymbol{\Theta} = (\Theta_{ij})_{1 \leq i, j \leq p} = \sum_{k=1}^r \bar{\lambda}_k \bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^T, \quad (2.3)$$

where $\bar{\lambda}_1 > \dots > \bar{\lambda}_r > 0$, and their corresponding eigenvectors are $\bar{\mathbf{q}}_1, \dots, \bar{\mathbf{q}}_r$. In this paper, we assume that the rank, r , of $\boldsymbol{\Theta}$ is fixed and finite. Similarly we decompose the integrated volatility matrix $\boldsymbol{\Gamma}$ as follows:

$$\mathbf{\Gamma} = (\Gamma_{ij})_{1 \leq i, j \leq p} = \sum_{k=1}^p \lambda_k \mathbf{q}_k \mathbf{q}_k^T, \quad (2.4)$$

where $\lambda_1 \cdots \lambda_p > 0$, and their corresponding eigenvectors are $\mathbf{q}_1, \dots, \mathbf{q}_p$.

Unfortunately we cannot observe the true log-prices due to the microstructural noises. In the high-frequency finance, we usually assume that observed log-price $Y_i(t_k)$ contains the true one contaminated with the additive noises:

$$Y_i(t_k) = X_i(t_k) + \varepsilon_i(t_k) \quad \text{for } i = 1, \dots, p, k = 0, \dots, n, \quad (2.5)$$

where $\varepsilon_i(t_k)$, $i = 1, \dots, p$, $k = 0, \dots, n$, are independent noises with mean zero, and for fixed i , $\varepsilon_i(t_k)$, $k = 0, \dots, n$, are i.i.d. random variables with variance η_{ii} and $\varepsilon_i(\cdot)$ is independent of the price processes $X_i(\cdot)$. For simplicity, we assume that the observed time points are regularly spaced, that is, $t_{k+1} - t_k = \pi^{-1}$, $k = 0, \dots, n-1$.

Remark 1

The assumption of the synchronized and equally spaced observation time points can be relaxed to the non-synchronized and unequally spaced observations. Then using some synchronized skill such as generalized sampling time (Aït-Sahalia et al., 2010), refresh time (Barndorff-Nielsen et al., 2011), and previous tick (Wang and Zou, 2010; Zhang, 2011) schemes, we can enjoy the same theoretical properties as we obtain in this paper. That is, the non-synchronization problem is manageable under some mild conditions (see Kim and Wang (2016); Tao et al. (2013a); Wang and Zou (2010)). So to focus on developing the robust volatility matrix estimator, we assume that the observation time points are synchronized and equally spaced.

3 Robust integrated volatility matrix estimation

In this section, we develop a robust volatility estimation procedure and show its concentration property with only finite fourth moments. First we introduce some notations. For any given p_1 by p_2 matrix $\mathbf{U} = (U_{ij})$, let

$$\|\mathbf{U}\|_1 = \max_{i \leq j \leq p_2} \sum_{i=1}^{p_1} |U_{ij}|, \quad \|\mathbf{U}\|_\infty = \max_{1 \leq i \leq p_1} \sum_{j=1}^{p_2} |U_{ij}|, \quad \|\mathbf{U}\|_{\max} = \max_{i,j} |U_{ij}|.$$

Matrix spectral norm $\|\mathbf{U}\|_2$ is the square root of the largest eigenvalue of $\mathbf{U}\mathbf{U}^T$. We denote the Frobenius norm of $\mathbf{U}$ by $\|\mathbf{U}\|_F = \sqrt{\text{tr}(\mathbf{U}^T \mathbf{U})}$, and define the relative Frobenius norm,

$$\|\mathbf{U}\|_{\mathbf{\Gamma}}^2 = p^{-1} \|\mathbf{\Gamma}^{-1/2} \mathbf{U} \mathbf{\Gamma}^{-1/2}\|_F^2. \quad (3.1)$$

Denote by C 's generic constants whose values are free of n and p and may change from appearance to appearance.

With bounded instantaneous volatilities and sub-Gaussian microstructural noises, Kim and Wang (2016) and Tao et al. (2013b) showed that some realized volatility matrix estimators have the sub-Gaussian concentration around the integrated volatility with convergence rate $n^{-1/4}$. However, it is well-known that the observed stock prices have heavy tails, and so the sub-Gaussian assumption is violated. Unfortunately, without the sub-Gaussian assumption, these types of realized volatility matrix estimators cannot achieve the sub-Gaussian concentration. Recently, Fan et al. (2016b,c) proposed a robust method for high-dimensional regression and covariance matrix by solving a Huber loss function and obtained the sub-Gaussian concentration with only finite fourth moments. Inspired by their robust M-estimator, we establish a robust estimation procedure as follows.

The observed log-returns, $Y_i(t_{k+1}) - Y_i(t_k)$, $i = 1, \dots, p$, $k = 0, \dots, n-1$, consist of the true log-return $X_i(t_{k+1}) - X_i(t_k)$ and microstructural noise $\varepsilon_i(t_{k+1}) - \varepsilon_i(t_k)$. Their volatilities are also coming from the true log-return volatility $\int_{t_k}^{t_{k+1}} \gamma_{ii}(t) dt$ and the variance of the microstructural noise $2\eta_{ii}$, which indicates that the variance of the noise dominates the volatility of the observed log-return. Thus, it is hard to investigate the volatility using the log-return observations. Hence, we first need to remove the microstructural noise by the following pre-averaging scheme:

$$Z_i(t_k) = \sum_{l=0}^{K_n-1} g\left(\frac{l}{K_n}\right) \{Y_i(t_{k+l+1}) - Y_i(t_{k+l})\} \quad \text{for } i = 1, \dots, p, k = 0, \dots, n - K_n, \quad (3.2)$$

where the weight function $g(\cdot)$ is continuous and piecewise continuous differentiable with a piecewise Lipschitz derivative g' and satisfies $g(0) = g(1) = 0$ and $\int_0^1 \{g(t)\}^2 dt > 0$. Then, with appropriate choice of diverging bandwidth parameter K_n , the volatilities of the pre-averaged data $Z_i(t_k)$, $i = 1, \dots, p$, $k = 0, \dots, n - K_n$, are no longer dominated by the variance of the microstructural noise. However, they are still heavy-tailed because of the heavy tails of the microstructural noises, and so we employ the Huber loss to mitigate the influence of the heavy tails as follows:

$$\hat{\theta}_{ij, \alpha} = \arg \min_{\theta} \sum_{k=0}^{n-K_n} \ell_{\alpha, ij} \left(\frac{n-K_n+1}{\phi K_n} Z_i(t_k) Z_j(t_k) - \theta \right), \quad (3.3)$$

where α_{ij} are truncation parameters and ℓ_{α} is the Huber loss

$$\ell_{\alpha}(x) = \begin{cases} 2\alpha|x| - \alpha^2 & \text{if } |x| \geq \alpha \\ x^2 & \text{if } |x| \leq \alpha, \end{cases}$$

and

$$\phi = \frac{1}{K_n} \sum_{l=0}^{K_n-1} \left\{ g\left(\frac{l}{K_n}\right) \right\}^2.$$

We call this robust estimator $\hat{\theta}_{ij,a}$ the robust pre-averaging realized volatility (RPRV) estimator. The RPRV estimator $\hat{\theta}_{ij,a}$ is the equivalent to the solution of the following equation

$$\sum_{k=0}^{n-K_n} \psi \left(\alpha_{ij}^{-1} \left(\frac{n-K_n+1}{\phi K_n} Z_{i(t_k)} Z_{j(t_k)} - \theta \right) \right) = 0,$$

where the influence function $\psi(x) = \text{sign}(x) \min(|x|, 1)$. If the tuning parameter α_{ij} is very big, the robust estimator will become the pre-averaging realized volatility estimator (Christensen et al., 2010; Jacod et al., 2009) and the bias is reduced. Therefore, it trades off between the biases and the heaviness of tails.

To investigate asymptotic properties of the RPRV estimator, we make some assumptions.

Assumption 1

- a. The drift processes are $\mu(t) = 0$ for $t \in [0, 1]$;
- b. There are some positive constants ν_γ and ν_e such that

$$\max_{1 \leq i \leq p} \gamma_{ii}(t) \leq \nu_\gamma \text{ a.s.}, \quad \max_{1 \leq i \leq p} E\{\varepsilon_{ii}^4(t_1)\} \leq \nu_e^2.$$

Remark 2—Assumption 1(a) is imposed for simplicity. The effects of the drifts $\mu(t)$ on realized volatility are higher order terms that are negligible, and they can be managed using appropriate decompositions as in Theorem 1 (Kim and Wang, 2016).

Remark 3—Assumption 1(b) indicates that the observed log-prices have only finite fourth moments due to the heavy-tailed microstructural noises. Then the existing realized volatility matrix estimators such as pre-averaging realized volatility (Christensen et al., 2010; Jacod et al., 2009) and multi-scale realized volatility (Zhang, 2006, 2011) estimators fail to obtain the sub-Gaussian concentration. On the other hand, if we investigate the asymptotic behaviors of volatility matrix estimators such as their convergence rate, we can relax the boundedness condition of the instantaneous volatility process $\gamma_{ii}(t)$ such as the locally bounded process (see Ait-Sahalia and Xiu (2015)) using the the localization argument made in Section 4.4.1 of Jacod and Protter (2012).

The theorem below shows that the RPRV estimator has the sub-Gaussian concentration.

Theorem 3.1

Suppose $K_n \log(\delta^{-1})/n = o(1)$ and $n/K_n^2 = O(1)$ for some $\delta > 0$. Take

$$\alpha_{ij} = \sqrt{\frac{(n - K_n + 1)\bar{V}_{ij}^*}{K_n \log \delta^{-1}}},$$

where

$$\bar{V}_{ij}^* = \frac{n - K_n + 1}{(\phi K_n)^2} \sum_{k=0}^{n - K_n} \text{var} \{Z_i(t_k)Z_j(t_k)\}.$$

Under models (2.1) and (2.5) and Assumption 1, we have for large n ,

$$\Pr \left\{ |\hat{\theta}_{ij,\alpha} - \theta_{ij,0}| \geq C \sqrt{\frac{(\nu_\gamma + \frac{n\zeta}{\phi K_n} \nu_\varepsilon)^4 K_n \log \delta^{-1}}{(n - K_n + 1)\bar{V}_{ij}^*}} + 2\nu_\gamma \frac{K_n}{n} \right\} \leq 3 \sqrt{\frac{n - K_n + 1}{K_n}} \delta, \quad (3.4)$$

where with $\eta_{ij} = 0$ for $i \neq j$,

$$\theta_{ij,0} = \Gamma_{ij} + \frac{n - K_n + 1}{\phi K_n} \zeta \eta_{ij}, \quad \zeta = \sum_{l=0}^{K_n - 1} \left\{ g\left(\frac{l}{K_n}\right) - g\left(\frac{l+1}{K_n}\right) \right\}^2 = O\left(\frac{1}{K_n}\right).$$

Remark 4—The term $\sqrt{(n - K_n + 1)/K_n}$ in (3.4) is coming from handling the randomness of the target parameter Γ_{ij} . Due to this term, the convergence rate of the estimators will be slightly slower up to $\log(p \vee \sqrt{n/K_n})$. Specifically, take $\delta = (p \vee \sqrt{n/K_n})^{-4}$ and we obtain

$$\Pr \left\{ \max_{1 \leq i, j \leq p} |\hat{\theta}_{ij,\alpha} - \theta_{ij,0}| \geq C \sqrt{\frac{K_n \log(p \vee \sqrt{n/K_n})}{n}} \right\} \leq p^{-1}.$$

When $p \propto n^c$ for some positive fixed constant c , the convergence rate is $\sqrt{K_n \log p/n}$ which is the optimal rate with $K_n = O(n^{1/2})$.

The robust estimator $\hat{\theta}_{ij,\alpha}$ in (3.3) is biased due to the microstructural noise, and the order of the bias term $(n - K_n + 1)\zeta \eta_{ij}/(K_n \phi)$ is n/K_n^2 . To make this asymptotically negligible, we can choose the tuning parameter $K_n = O(n^{3/5})$, and then the convergence rate will be $n^{-1/5}$, which is not optimal. To achieve the optimal convergence rate $n^{-1/4}$, we need to correct the bias. First, we estimate the variance of the microstructural noise, η_{ij} , by minimizing the Huber loss

$$\hat{\eta}_{ii} = \arg \min_{\theta} \sum_{k=0}^{n-1} \ell_{\alpha_{\eta}} [\{Y_i(t_{k+1}) - Y_i(t_k)\}^2/2 - \theta], \quad (3.5)$$

where the tuning parameter $\alpha_{\eta,ii}$ will be given in Theorem 3.2. Then, define the estimator of the integrated volatility Γ_{ij} as

$$\hat{\Gamma}_{ij} = \hat{\theta}_{ij,\alpha} - \frac{n - K_n + 1}{\phi K_n} \zeta \hat{\eta}_{ij} \mathbf{1}(i = j),$$

where $\hat{\theta}_{ij,\alpha}$ is the solution of (3.3) with $K_n = C_K n^{1/2}$ and $\alpha_{ij} = \sqrt{(n - K_n + 1) \bar{V}_{ij}^* / (K_n \log \delta^{-1})}$ for some constant C_K free of n and p .

Denote by $\hat{\mathbf{\Gamma}} = (\hat{\Gamma}_{ij})_{1 \leq i,j \leq p}$ the robust pre-averaging realized volatility matrix (RPRVM) estimator. The following theorem shows that the RPRVM estimator $\hat{\mathbf{\Gamma}}$ possesses the sub-Gaussian concentration.

Theorem 3.2

Under the assumptions in Theorem 3.1, take

$$\alpha_{\eta,ii} = \sqrt{\frac{n \bar{V}_{\varepsilon,ii}^*}{2 \log \delta^{-1}}}, \quad \text{where } \bar{V}_{\varepsilon,ii}^* = \frac{1}{4n} \sum_{k=0}^{n-1} \text{var} [\{Y_i(t_{k+1}) - Y_i(t_k)\}^2].$$

Then we have for large n ,

$$\Pr \left\{ |\hat{\eta}_{ii} - \eta_{ii}| \geq C \sqrt{\frac{\nu_{\varepsilon}^4 \log \delta^{-1}}{n \bar{V}_{\varepsilon,ii}^*}} + \frac{1}{2n} \nu_{\gamma} \right\} \leq 2\sqrt{n}\delta$$

and

$$\Pr \left\{ |\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq C \sqrt{\frac{(\nu_{\gamma} + \frac{n \zeta}{\phi K_n} \nu_{\varepsilon})^4 K_n \log \delta^{-1}}{(n - K_n + 1) \bar{V}_{ij}^*}} + \frac{2\nu_{\gamma} K_n}{n} \right\} \leq 4 \sqrt{\frac{n - K_n + 1}{K_n}} \delta. \quad (3.6)$$

Remark 5—Theorem 3.2 shows that the robust volatility estimator $\hat{\Gamma}_{ij}$ has the sub-Gaussian concentration with the convergence rate $n^{-1/4}$ which is optimal.

Throughout the paper, we develop the large volatility matrix estimator based on the RPRVM estimator $\hat{\mathbf{\Gamma}} = (\hat{\Gamma}_{ij})_{1 \leq i,j \leq p}$.

4 Large volatility matrix estimation via factor models

To account for the common factors in the financial market, as discussed in Section 2, the integrated volatility matrix has two components:

$$\Gamma = \sum_{k=1}^p \lambda_k \mathbf{q}_k \mathbf{q}_k^\top = \Theta + \Sigma$$

The low-rank volatility matrix Θ comes from the common market factors and represents the systematic risk. The remaining volatilities stand for the specific risk (or idiosyncratic risk), which is specific to individual firms. Thus, their volatility matrix Σ should be sparse (Fan et al., 2013, 2016a; Kim et al., 2015b). In the light of this, we assume that Σ satisfies the sparse condition,

$$\max_{1 \leq i \leq p} \sum_{j=1}^p |\Sigma_{ij}|^q (\sum_{ii} \sum_{jj})^{(1-q)/2} \leq M_\sigma s_p, \quad (4.1)$$

where M_σ is a positive random variable with $E(M_\sigma^2) < \infty$, $q \in [0, 1)$, and s_p is a deterministic function of p that grows slowly in p , representing the sparsity level of the volatility matrix. For example, when Σ_{ij} is bounded from below and $q = 0$, s_p measures the maximum number of nonvanishing elements in each row of the matrix Σ . Therefore, the integrated volatility matrix Γ has the low-rank plus sparse structure, which is widely used in analyzing the large covariance matrix (Fan et al., 2013, 2016c) and large integrated volatility matrix (Ait-Sahalia and Xiu, 2015; Fan et al., 2016a; Kim et al., 2015b).

Remark 6

There is an identifiability issue in a latent factor model. Hence, in the literature of approximate factor models (Doz et al., 2011; Fan et al., 2013), some identification restrictions are often employed. For example, the low-rank volatility matrix Θ is assumed to have the decomposition, $\Theta = \mathbf{L}\mathbf{L}^\top$, and the factor loading matrix $\mathbf{L}$ to satisfy that $\mathbf{L}^\top \mathbf{L}$ is diagonal. However, in this paper, our main purpose is to estimate the large integrated volatility matrix consisting of the low-rank volatility and sparse volatility matrices. So we do not impose these identification restrictions, but we still need to separate the low-rank volatility matrix Θ and sparse volatility matrix Σ from the integrated volatility matrix Γ . As discussed in Ait-Sahalia and Xiu (2015), these two matrices are well-identifiable under the incoherence and pervasive conditions as long as $s_p/p^{1/2} = o(1)$.

Fan et al. (2013) introduced the principal orthogonal component thresholding (POET) procedure to estimate large covariance matrices based on the approximate factor model. We employ the POET procedure to estimate the large integrated volatility matrix consisting of the low-rank and sparse matrices. The POET procedure follows three steps:

1. Using the eigen-decomposition, we can decompose the robust estimator $\hat{\Gamma}$ as follows:

$$\hat{\Gamma} = \sum_{k=1}^p \hat{\lambda}_k \hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^{\top},$$

where $\hat{\lambda}_1 \cdots \hat{\lambda}_p$ and $\hat{\mathbf{q}}_k, k = 1, \dots, p$, are their corresponding eigenvectors. We estimate Θ by $\hat{\Theta} = \sum_{k=1}^r \hat{\lambda}_k \hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^{\top}$.

2. The sparse volatility matrix Σ is estimated by

$$\hat{\Sigma}_{ij} = \begin{cases} \tilde{\Sigma}_{ij} \vee 0, & \text{if } i = j \\ s_{ij}(\tilde{\Sigma}_{ij}) \mathbf{1}(|\tilde{\Sigma}_{ij}| \geq \varpi_{ij}), & \text{if } i \neq j, \end{cases}$$

where $\tilde{\Sigma} = (\tilde{\Sigma}_{ij})_{1 \leq i, j \leq p} = \hat{\Gamma} - \hat{\Theta}$, $s_{ij}(\cdot)$ satisfies that $|s_{ij}(x) - x| \leq \varpi_{ij}$ and the adaptive thresholding level $\varpi_{ij} = \varpi_n \sqrt{(\sum_{ii} \tilde{\Sigma} \vee 0)(\sum_{jj} \tilde{\Sigma} \vee 0)}$, applied to the correlation matrix. Interesting examples of thresholding function $s_{ij}(x)$ include the hard thresholding function $s_{ij}(x) = x$ and the soft thresholding function $s_{ij}(x) = x - \text{sign}(x) \varpi_{ij}$. The tuning parameter ϖ_n will be determined in Theorem 4.1.

3. With the sparse volatility matrix estimator $\hat{\Sigma} = (\hat{\Sigma}_{ij})_{1 \leq i, j \leq p}$, we estimate the integrated volatility matrix Γ by

$$\hat{\Gamma}_{POET} = \sum_{k=1}^r \hat{\lambda}_k \hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^{\top} + \hat{\Sigma}.$$

To investigate the robust volatility matrix estimator $\hat{\Gamma}_{POET}$, we make several conditions.

Assumption 2

- a. For some fixed constant C_1 , we have

$$\frac{p}{r} \max_{1 \leq i \leq p} \sum_{j=1}^r \bar{q}_{ij}^2 \leq C_1 \text{ a.s.},$$

where $\bar{\mathbf{q}}_j = (\bar{q}_{1j}, \dots, \bar{q}_{pj})^{\top}$ is the j -th eigenvector of Θ ;

- b. Let $D_{\lambda} = \min\{\bar{\lambda}_i - \bar{\lambda}_{i+1} : 1 \leq i \leq r\}$, and $(\lambda_1 + pM_{\sigma})/D_{\lambda} \leq C_2$ a.s. and $D_{\lambda} \geq C_3 p$ a.s. for some generic constants C_2 and C_3 , where $\bar{\lambda}_{r+1} = 0$, M_{σ} is defined in (4.1), and $\bar{\lambda}_i$ and λ_i are the i -th eigenvalues of Θ and Γ , respectively;
- c. The smallest eigenvalue of Σ stays away from zero almost surely;
- d. $s_p/\sqrt{p} + \sqrt{\log p/n^{1/2}} = o(1)$.

Remark 7—Assumption 2(a) is called the incoherence condition which is widely used in analyzing low-rank matrices (see Candès et al. (2011); Fan et al. (2016c)). Under the incoherence condition, Assumption 2(b) is usually satisfied under some mild condition for the factor loading matrix (see Proposition 3.2 in Fan et al. (2016c)).

The following theorem shows the asymptotic behaviors of POET estimators.

Theorem 4.1

Under the model (2.1), suppose that the concentration inequality,

$$\Pr \left\{ \max_{1 \leq i, j \leq p} |\hat{\Gamma}_{ij} - \Gamma_{ij}| \geq C \sqrt{\frac{\log(p \vee n)}{n^{1/2}}} \right\} \leq p^{-1}, \quad (4.2)$$

Assumption 2, and sparsity condition (4.1) are met. Take $\varpi_n = C\varpi\beta_n$ for some large fixed constant $C\varpi$, where $\beta_n = M_{\sigma^2 p}/p + \sqrt{\log(p \vee n)/n^{1/2}}$. Then with probability greater than $1 - 2p^{-1}$, the POET estimator satisfies

$$\|\hat{\Sigma} - \Sigma\|_2 \leq CM_{\sigma^2 p}\beta_n^{1-q}, \quad (4.3)$$

$$\|\hat{\Sigma} - \Sigma\|_{\max} \leq C\beta_n, \quad (4.4)$$

$$\|\hat{\Gamma}_{POET} - \Gamma\|_{\Gamma} \leq C \left(\frac{p^{1/2} \log(p \vee n)}{n^{1/2}} + M_{\sigma^2 p}\beta_n^{1-q} \right), \quad (4.5)$$

$$\|\hat{\Gamma}_{POET} - \Gamma\|_{\max} \leq C\beta_n, \quad (4.6)$$

where the relative Frobenius norm $\|\cdot\|_{\Gamma}$ is defined in (3.1). Furthermore, if $M_{\sigma^2 p}\beta_n^{1-q} = o(1)$, then, with probability approaching 1, the eigenvalues of $\hat{\Sigma}$ are all greater than 0, $\hat{\Gamma}_{POET}$ is non-singular, and

$$\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_2 \leq CM_{\sigma^2 p}\beta_n^{1-q}, \quad (4.7)$$

$$\|\hat{\Gamma}_{POET}^{-1} - \Gamma^{-1}\|_2 \leq CM_{\sigma^2} p \beta_n^{1-q}. \quad (4.8)$$

Remark 8—Theorem 4.1 shows that for any initial volatility matrix estimators satisfying the concentration condition (4.2), its resulting POET estimator enjoys the above asymptotic properties. For example, the RPRVM estimator with $\delta = (p \vee n^{1/4})^{-4}$ satisfies (4.2).

Remark 9—Theorem 4.1 indicates that the POET estimator can still be consistent as long as $p = o(n)$ in terms of the relative Frobenius norm, which is comparable with the convergence rate of the blocked and regularized volatility matrix (BRVM) estimator in Kim et al. (2015b). On the other hand, the POET estimator for large covariance matrices (Fan et al., 2013, 2016c) has the convergence rates with n instead of $n^{1/2}$ in the convergence rates obtained in Theorem 4.1. This difference may come from the microstructural noise contamination. That is, if the true stock prices are directly observable, the convergence rates will become the same as those in Fan et al. (2013, 2016c).

4.1 Discussion on the tuning parameter selection

In practice, it is important but difficult to choose tuning parameters. In this section, we discuss how to develop a data driven tuning parameter selection method. First, we consider the bandwidth K_n and weight (kernel) function $g(\cdot)$ for the pre-averaging scheme. We can choose K_n and $g(\cdot)$ which minimize the asymptotic variance of the pre-averaging realized volatility estimator (Christensen et al., 2010; Jacod et al., 2009). To do this, we can employ the asymptotic variance estimator proposed by Kim et al. (2015a),

$$\hat{v}_{ij} = \frac{2}{\phi^2} \left\{ \phi_1 C_K(\hat{\kappa}_{1,ij} + \hat{\kappa}_{2,ij}) + \frac{\phi_2}{C_K} \left(\hat{\Gamma}_{ii}^{PR} \hat{\eta}_{jj}^{PR} + \hat{\Gamma}_{jj}^{PR} \hat{\eta}_{ii}^{PR} \right) + \frac{\phi_3}{C_K^3} \hat{\eta}_{ii}^{PR} \hat{\eta}_{jj}^{PR} \right\}, \quad (4.9)$$

where

$$\hat{\Gamma}_{ij}^{PR} = \frac{1}{\phi K_n} \sum_{k=0}^{n-K_n} [Z_i(t_k) Z_j(t_k) - \hat{\eta}_{ij}^{PR} \mathbf{1}(i=j)], \quad (4.10)$$

$$\hat{\eta}_{ii}^{PR} = \frac{1}{2n} \sum_{k=0}^{n-1} [Y_i(t_{k+1}) - Y_i(t_k)]^2, \quad \hat{\kappa}_{1,ij} = \frac{1}{\phi^2 K_1^2 n^{-1}} \sum_{k=0}^{n-2K_1} Z_i(t_k)^2 Z_j(t_{k+K_1})^2, \quad (4.11)$$

$$\hat{\kappa}_{2,ij} = \frac{1}{\phi^2 K_1^{2n-1}} \sum_{k=0}^{n-2K_1} Z_i(t_k) Z_j(t_k) Z_i(t_k + K_1) Z_j(t_k + K_1),$$

$$\phi_1 = \int_0^1 \left[\int_s^1 g(u)g(u-s)du \right]^2 ds,$$

$$\phi_2 = \int_0^1 \left[\int_s^1 g'(u)g'(u-s)du \int_s^1 g(t)g(t-s)dt \right] ds,$$

$$\phi_3 = \int_0^1 \left[\int_s^1 g'(u)g'(u-s)du \right]^2 ds, \quad K_1 = \lfloor n^{2/3} \rfloor,$$

and the pre-averaged observation $Z_A(t_k)$ is defined in (3.2). To evaluate the asymptotic variance estimator $\hat{\nu}_{j\hat{p}}$, we can choose initially $K_n = \lfloor n^{1/2} \rfloor$. Then, for given weight function $g(\cdot)$, we can find the optimal bandwidth size $K_n = C_K n^{1/2}$. On the other hand, it is difficult to find the general optimal weight (kernel) function over the class satisfying given conditions in (3.2). This may require some further extensive fine analysis and we do not expect the results to depend sensitively on the weight function. We leave this issue for the future study. In this paper, we choose the weight function $g(x) = x \wedge (1-x)$ for the numerical studies.

Next, we consider the choice of the rank r . Under the model set-up such as the pervasive condition (Assumption 2(b)) and sparse condition (4.1), the r -th eigenvalue diverges with the order p while the $r+1$ -th eigenvalue diverges with at most order s_p . Based on this fact, we can develop some method to choose the rank r . For example, Ait-Sahalia and Xiu (2015) suggested the following method

$$\hat{r} = \arg \min_{1 \leq j \leq r_{\max}} \left(p^{-1} \hat{\lambda}_j + j \times f(n, p) \right) - 1,$$

where the penalty function $f(n, p)$ satisfies $f(n, p) \rightarrow 0$ and

$f(n, p) \times (\sqrt{\log p/n^{1/2}} + p^{-1}s_p)^{-1} \rightarrow \infty$, for example, $f(n, p) = c_1 (\sqrt{\log p/n^{1/2}} + p^{-1}s_p)^{c_2}$ for some fixed constants c_1, c_2 , and $0 < c_2 < 1$. Then, under some mild condition, they showed that $\Pr(\hat{r} = r) \rightarrow 1$ (see Theorem 2 in Ait-Sahalia and Xiu (2015)). There are some alternative methods to determine the number of factors (Ahn and Horenstein, 2013; Alessi et al., 2010; Bai and Ng, 2002; Fan et al., 2013; Hallin and Liška, 2007; Lam and Yao, 2012). It is worthwhile to mention that an over estimate of $\hat{r}$ may not do as much harm to estimating the integrated volatility matrix Γ . On the other hand, an underestimate of $\hat{r}$ can result in the inconsistency of the POET estimator because of missing important factors (Fan et al., 2013).

Finally, we consider the selection of thresholding level ϖ_n . We can use some cross validation idea on multi-period high-frequency financial data such as high-frequency data over many days. For example, we impose the usual realistic stationarity assumption on daily sparse volatility matrix in financial time series and adopted one-day-ahead prediction error as a performance evaluation measure for threshold selection (Wang and Zou, 2010). We define the mean squared prediction error (MSPE) as follows:

$$MSPE\{\mathcal{T}_{\varpi}(\tilde{\Sigma})\} = \frac{1}{d} \sum_{k=1}^{d-1} \|\mathcal{T}_{\varpi}(\tilde{\Sigma}_k) - \tilde{\Sigma}_{k+1}\|_F^2, \quad (4.12)$$

where ϖ is the tuning parameter, $\tilde{\Sigma}_k = \hat{\Gamma}_k - \hat{\Theta}_k$ is a preliminary estimators for the sparse matrix, $\hat{\Theta}_k$ is low-rank matrix estimators, and the thresholding function is

$$\mathcal{T}_{\varpi}(\tilde{\Sigma}) = \left(\left(\tilde{\Sigma}_{ij} - \text{sign}(\tilde{\Sigma}_{ij})\varpi_{ij} \right) \mathbf{1}(|\tilde{\Sigma}_{ij}| \geq \varpi_{ij}, i \neq j) \right)_{1 \leq i, j \leq p} + \left((\tilde{\Sigma}_{ii} \vee 0) \mathbf{1}(i = j) \right)_{1 \leq i, j \leq p}$$

with $\varpi_{ij} = \varpi \sqrt{(\tilde{\Sigma}_{ii} \vee 0)(\tilde{\Sigma}_{jj} \vee 0)}$. The thresholding level can be found by minimizing MSPE.

5 A simulation study

We conducted a simulation study to check the finite sample performance of the POET estimator constructed using the robust realized volatility matrix estimator as an input. The true log-prices follow the model (2.1) with $\mu(t) = 0$. We used two types of instantaneous volatility processes; the heavy tail and sub-Gaussian processes. The heavy tail process is generated as follows. Let $\sigma(t)$ be the Cholesky decomposition of $\zeta(t) = (\zeta_{ij}(t))_{1 \leq i, j \leq p}$. The diagonal elements of $\zeta(t)$ come from four different processes which are geometric Ornstein-Uhlenbeck processes, the sum of two CIR processes (Cox et al., 1985; Barndorff-Nielsen, 2002), the volatility process in Nelson's GARCH diffusion limit model (Wang, 2002), and two-factor log-linear stochastic volatility process (Huang and Tauchen, 2005) with leverage effect. Details can be found in Wang and Zou (2010). To control the tail behaviors of the instantaneous volatility matrix $\zeta(t)$, we used the t-distribution as follows:

$$\varsigma_{ii}(t_k) = (1 + t_{5,k}^2) \tilde{\varsigma}_{ii}(t_k),$$

where $t_{df,k}$, $k = 1, \dots, n$, are the i.i.d. t-distributions with degrees of freedom df , and $\tilde{\varsigma}_{ii}(t_k)$, $k = 1, \dots, n$, were generated in the same way in Wang and Zou (2010). To obtain the sparse instantaneous matrix $\zeta(t)$, its off-diagonal elements were generated as follows:

$$\varsigma_{ij}(t_k) = \{\kappa(t_k)\}^{|i-j|} \sqrt{\varsigma_{ii}(t_k)\varsigma_{jj}(t_k)}, \quad 1 \leq i \neq j \leq p,$$

where the process $\kappa(t)$ is given by

$$\kappa(t) = \frac{e^{\frac{1}{2}u(t)} - 1}{e^{\frac{1}{2}u(t)} + 1}, \quad du(t) = 0.03\{0.64 - u(t)\}dt + 0.118u(t)dW_{\kappa,t},$$

$$W_{\kappa,t} = \sqrt{0.96}W_{\kappa,t}^0 - 0.2 \sum_{i=1}^p W_{it}/\sqrt{p},$$

and $W_{\kappa,t}^0$, $\kappa = 1, \dots, p$, are one-dimensional Brownian motions independent of the Brownian motions W_t^* and W_t . The low-rank instantaneous volatility matrix $\boldsymbol{\vartheta}^\top(t)\boldsymbol{\vartheta}(t) = \boldsymbol{\zeta}^f(t)$ is $\mathbf{L}^\top \{\boldsymbol{\vartheta}^f(t)\}^\top \boldsymbol{\vartheta}^f(t) \mathbf{L}$, where $\mathbf{L} = (L_{ij})_{1 \leq i \leq r, 1 \leq j \leq p} \in \mathbb{R}^{r \times p}$, L_{ij} were generated from the i.i.d. uniform distributions on $[-2, 2]$, and $\boldsymbol{\vartheta}^f(t)$ was generated similar to $\boldsymbol{\sigma}(t)$. Specifically, $\boldsymbol{\vartheta}^f(t)$ is the Cholesky decomposition of $\boldsymbol{\zeta}^f(t)$, and the diagonal elements of $\boldsymbol{\zeta}^f(t)$ at observation time t_k are

$$\varsigma_{ii}^f(t_k) = \left\{ 1 + (t_{5,k}^f)^2 \right\} \tilde{\varsigma}_{ii}^f(t_k),$$

where $t_{df,k}^f$, $k = 1, \dots, n$, are the i.i.d. t-distribution with degrees of freedom df , and $\tilde{\varsigma}_{ii}^f(t_k)$, $k = 1, \dots, n$, were generated from the three different processes (geometric Ornstein-Uhlenbeck processes, the sum of two CIR processes (Cox et al., 1985; Barndorff-Nielsen, 2002), and the volatility process in Nelson's GARCH diffusion limit model (Wang, 2002)). Its off-diagonal elements were set as zero.

Secondly, to generate the sub-Gaussian process, we used constant instantaneous volatilities. The sparse instantaneous volatility matrix $\boldsymbol{\zeta}(t)$ is

$$\varsigma_{ij}(t) = 0.4^{|i-j|} \sqrt{\varsigma_{ii}(t)\varsigma_{jj}(t)}, \quad 1 \leq i \neq j \leq p,$$

where $\varsigma_{ii}(t) = 10$ for $i = 1, \dots, p$. The low-rank matrix was generated similar to the heavy tail process except for instantaneous volatilities $\varsigma_{ii}^f(t) = 1$, $i = 1, \dots, r$.

Finally, the high-frequency observations $Y_{\lambda}(t_k)$ were generated from model (2.5) with noise $\varepsilon_{\lambda}(t_k)$ obtaining from independent t-distribution with degrees of freedom df and standard deviation

$$\sqrt{\eta_{ii}} = 0.01 \sqrt{\int_0^1 r_{ii}(t) dt},$$

where $df = 5$ for the heavy tail process and $df = \infty$ for the sub-Gaussian process.

We chose $p = 512$ and $r = 3$, and the number, n , of observations was varied from 250 to 1000. We used the simulated observations $Y_i(t_k)$ to calculate the POET estimators using the pre-averaging realized volatility matrix (PRVM) estimator (Christensen et al., 2010; Jacod et al., 2009) and proposed robust estimator, RPRVM, defined in (3.3). The PRVM estimator is calculated by $(\hat{\Gamma}_{ij}^{PR})_{1 \leq i, j \leq p}$ in (4.10). We used the weight function $g(x) = x \wedge (1 - x)$ and $K_n = \lfloor n^{1/2} \rfloor$ for the PRVM and RPRVM estimators. For the RPRVM estimator, we used the following tuning parameter,

$$\alpha_{ij} = \sqrt{\frac{(n - K_n + 1)\hat{v}_{ij}}{K_n \log p}}, \quad (5.1)$$

where the asymptotic variance estimator, $\hat{v}_{ij}$, is given in (4.9). Similarly, we used the heterogeneous tuning parameters α_η 's,

$$\alpha_{\eta, ii} = \sqrt{\frac{n\hat{v}_{\eta, ii}}{2 \log p}}, \quad (5.2)$$

where the asymptotic variance of the estimator of η_{ii} is given by

$$\hat{v}_{\eta, ii} = \frac{1}{n} \sum_{k=0}^{n-1} [\{Y_i(t_{k+1}) - Y_i(t_k)\}^2/2 - \hat{\eta}_{ii}^{PR}]^2,$$

where $\hat{\eta}_{ii}^{PR}$ is defined in (4.11). For the thresholding step, we used the soft thresholding scheme and selected the threshold level by minimizing the corresponding Frobenius norm. Unfortunately, for the finite sample, it is not guaranteed that RPRVM and PRVM estimators are positive semi-definite (PSD). To investigate the impact of the positive semi-definiteness, we projected the RPRVM and PRVM estimators onto the positive semi-definite cone in the spectral norm. Specifically, we replace the negative eigenvalues of the RPRVM and PRVM estimators by zero, and the eigenvectors remain the same as those of the RPRVM and PRVM estimators. We denote these positive semi-definite version of the RPRVM and PRVM estimators by PRPRVM and PPRVM estimators, respectively. The ℓ_2 -norms (spectral norms) and the relative Frobenius norms, $\|\cdot\|_F$, of the difference between estimator and true integrated volatility matrix were computed based on 100 simulations.

Figure 1 plots the relative Frobenius norm errors against the sample size n for RPRVM, PRVM, PRPRVM, and PPRVM estimators and corresponding thresholding estimators and POET estimators, and Table 1 reports their numerical results. We find that the proposed robust estimators, the RPRVM and PRPRVM estimators, show better performance than the PRVM and PPRVM estimators in terms of the relative Frobenius norm error even for the sub-Gaussian case. The POET estimators outperform the thresholding and original realized volatility estimators (PRVM, RPRVM, PPRVM, and PRPRVM), but the thresholding scheme cannot improve the performance significantly due to the nonsparse nature of the

integrated volatility matrix Γ . When comparing the positive semi-definite estimators and original volatility estimators, there are no significant difference. In fact, the integrated volatility matrix can be explained by first few leading eigen factors, and so even though there are some negative eigenvalues, they cannot affect significantly on estimating Γ .

Figures 2–3 depicts the spectral norm errors against the sample size n for the estimators of the sparse matrix Σ and inverse matrix Γ^{-1} calculated using the POET procedures with the RPRVM, PRVM, PRPRVM, and PPRVM estimators. We find that in most of the cases, the projected estimators improve the performances of estimating the sparse volatility and inverse volatility matrices. The reason is that these matrices are affected by small eigenvalues. After projecting the estimators onto the positive semi-definite space, the robust estimators, the RPRVM estimators, show better performance than PPRVM estimators when the observations have the heavy tail process. On the other hand, for the sub-Gaussian process, the RPRVM estimators cannot outperform against the PPRVM estimators especially for estimating sparse volatility matrices.

6 An empirical study

In this section, we applied the POET estimators to high-frequency trading data for 200 assets over a period of 63 days in 2013. We selected top 200 large trading volume stocks among S&P 500 and used 1-min log-returns. The data is taken from the Wharton Data Service (WRDS) system. To evaluate POET estimators, we need to determine the rank r . First we calculated 63 daily integrated volatilities using the RPRVM and PRVM estimation procedures. Then we used the scree plot and draw it using the median values of eigenvalues from 63 RPRVM and PRVM estimators. Figure 4 shows that $r = 1, 2, 3, 4, 5$ are the possible values of the number of factors.

We employed one-day-ahead prediction error defined in Section 4.1 as a performance measure for evaluations. Denote by $\hat{\Gamma}_k$, $k = 1, \dots, 63$, the realized volatility matrix estimators being one of RPRVM and PRVM estimators. The thresholding levels were determined by minimizing the MSPE in (4.12) with $d = 63$. We chose the tuning parameters $K_n = \lfloor n^{1/2} \rfloor$ and $K_1 = \lfloor n^{3/5} \rfloor$ for RPRVM and PRVM estimators, and the tuning parameters α 's in (5.1) and (5.2) were used for the Huber loss function.

Table 2 reports MSPE results for sparse volatility estimators calculated from the POET procedures with the RPRVM and PRVM estimators. We find that when the rank, r , is one, both of RPRVM and PRVM estimators show the best performance. When comparing the RPRVM and PRVM estimators, the robust estimator, the RPRVM estimator, has smaller MSPE. These findings indicate that the integrated volatility matrices may have some factor structure with rank 1, and the stock prices have heavy tails.

7 Proofs

7.1 Proofs of Section 3

Proof of Theorem 3.1—To ease the notations, we omit the subscript n , for example, K . Without loss of generality, we assume that $n = K(L + 1) - 1$ for some $L \in \mathbb{N}$. Let

$$f(\theta) = \frac{\alpha_{ij}}{(n-K+1)} \sum_{k=0}^{n-K} \psi\{\alpha_{ij}^{-1}(Q_{ij}(t_k) - \theta)\},$$

$$A_{ij}(t_k) = \frac{n-K+1}{\phi K} \left[\sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) \right\}^2 E \left\{ \int_{t_k+l}^{t_k+l+1} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right\} + \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) - g\left(\frac{l+1}{K}\right) \right\}^2 \eta_{ij} \right],$$

$$V_{ij}(t_k) = E[\{Q_{ij}(t_k) - A_{ij}(t_k)\}^2 | \mathcal{F}_{t_k}],$$

where

$$Q_{ij}(t_k) = \frac{n-K+1}{\phi K} Z_i(t_k) Z_j(t_k).$$

Simple algebra shows

$$\begin{aligned} Q_{ij}(t_k) - A_{ij}(t_k) &= \frac{n-K+1}{\phi K} \bar{X}_i(t_k) \bar{X}_j(t_k) - \frac{n-K+1}{\phi K} \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) \right\}^2 E \left\{ \int_{t_k+l}^{t_k+l+1} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right\} \\ &\quad + \frac{n-K+1}{\phi K} \bar{\varepsilon}_i(t_k) \bar{\varepsilon}_j(t_k) - \frac{n-K+1}{\phi K} \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) - g\left(\frac{l+1}{K}\right) \right\}^2 \eta_{ij} + \frac{n-K+1}{\phi K} \bar{X}_i(t_k) \bar{\varepsilon}_j(t_k) + \frac{n-K+1}{\phi K} \bar{\varepsilon}_i(t_k) \bar{X}_j(t_k) \\ &= (a) + (b) + (c) + (d), \end{aligned}$$

where $\eta_{ij} = 0$ for $i \neq j$,

$$\bar{\varepsilon}_i(t_k) = \sum_{l=0}^{K-1} g\left(\frac{l}{K}\right) \{\varepsilon_i(t_{k+l+1}) - \varepsilon_i(t_{k+l})\} = \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) - g\left(\frac{l+1}{K}\right) \right\} \varepsilon_i(t_{k+l+1})$$

and

$$\bar{X}_i(t_k) = \sum_{l=0}^{K-1} g\left(\frac{l}{K}\right) \{X_i(t_{k+l+1}) - X_i(t_{k+l})\}.$$

Since ε and X are independent,

$$E\{(c) + (d) | \mathcal{F}_{t_k}\} = 0 \text{ a.s.}$$

We can easily show

$$E\{(b)|\mathcal{F}_{t_k}\} = 0 \text{ a.s.}$$

Finally, by the application of Itô's isometry, we can show

$$\begin{aligned} & E \left[\hat{X}_{i(t_k)} \hat{X}_{j(t_k)} - \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) \right\}^2 \int_{t_k+l}^{t_k+l+1} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right] \\ &= E \left[\sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) \right\}^2 \left[\{X_{i(t_k+l+1)} - X_{i(t_k+l)}\} \{X_{j(t_k+l+1)} - X_{j(t_k+l)}\} \right. \right. \\ &\quad \left. \left. - E \left[\int_{t_k+l}^{t_k+l+1} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right] \right] \middle| \mathcal{F}_{t_k} \right] + E \left[\sum_{l \neq l'} g\left(\frac{l}{K}\right) g\left(\frac{l'}{K}\right) \{X_{i(t_k+l+1)} - X_{i(t_k+l)}\} \{X_{j(t_k+l'+1)} - X_{j(t_k+l')} \} \right. \\ &\quad \left. \middle| \mathcal{F}_{t_k} \right] = 0 \text{ a.s.} \end{aligned}$$

Thus,

$$E\{Q_{ij}(t_k) - A_{ij}(t_k) | \mathcal{F}_{t_k}\} = 0 \text{ a.s.} \quad (7.1)$$

Similar to the proofs of Theorem 1 (Kim and Wang, 2016), we can show

$$E[\{Q_{ij}(t_k) - A_{ij}(t_k)\}^2 | \mathcal{F}_{t_k}] \leq C \left(\nu_\gamma^2 + \frac{n^2 \zeta^2}{\phi^2 K^2} \nu_\varepsilon^2 + \frac{3\zeta n}{\phi K} \nu_\gamma \nu_\varepsilon \right) \text{ a.s.}, \quad (7.2)$$

which is finite.

Define

$$\begin{aligned} B_+(\theta) &= \frac{1}{n-K+1} \sum_{k=0}^{n-K} \left[A_{ij}(t_k) - \theta + \alpha_{ij}^{-1} \{V_{ij}(t_k) + (A_{ij}(t_k) - \theta)^2\} + \frac{\log(1/\delta)}{L\alpha_{ij}^{-1}} \right], \\ B_-(\theta) &= \frac{1}{n-K+1} \sum_{k=0}^{n-K} \left[A_{ij}(t_k) - \theta - \alpha_{ij}^{-1} \{V_{ij}(t_k) + (A_{ij}(t_k) - \theta)^2\} - \frac{\log(1/\delta)}{L\alpha_{ij}^{-1}} \right]. \end{aligned}$$

Let

$$B_{ij}(t_k) = A_{ij}(t_k) - \theta + \alpha_{ij}^{-1} \{V_{ij}(t_k) + (A_{ij}(t_k) - \theta)^2\} + \frac{\log(1/\delta)}{L\alpha_{ij}^{-1}}.$$

We have

$$\begin{aligned} & \Pr \{f(\theta) > B_+(\theta)\} \\ &= \Pr \left[\exp\{\alpha_{ij}^{-1} \frac{n-K+1}{K} f(\theta)\} > \exp\{\alpha_{ij}^{-1} \frac{n-K+1}{K} B_+(\theta)\} \right] \\ &\leq \mathbb{E} \left[\mathbf{1} \left(\exp\{\alpha_{ij}^{-1} \frac{n-K+1}{K} f(\theta)\} > \exp\{\alpha_{ij}^{-1} \frac{n-K+1}{K} B_+(\theta)\} \right) \right] \\ &\leq \mathbb{E} \left[\frac{\exp\{\alpha_{ij}^{-1} \frac{n-K+1}{K} f(\theta)\}}{\exp\{\alpha_{ij}^{-1} \frac{n-K+1}{K} B_+(\theta)\}} \right] \\ &= \mathbb{E} \left[\prod_{m=0}^{K-1} \prod_{k=0}^{L-1} \frac{\exp[K^{-1} \psi\{\alpha_{ij}^{-1}(Q_{ij}(t_{Kk+m}) - \theta)\}]}{\exp\{K^{-1} \alpha_{ij}^{-1} B_{ij}(t_{Kk+m})\}} \right] \\ &\leq \prod_{m=0}^{K-1} \mathbb{E} \left[\prod_{k=0}^{L-1} \frac{\exp[\psi\{\alpha_{ij}^{-1}(Q_{ij}(t_{Kk+m}) - \theta)\}]}{\exp\{\alpha_{ij}^{-1} B_{ij}(t_{Kk+m})\}} \right]^{1/K} \\ &= \prod_{m=0}^{K-1} \mathbb{E} \left[\prod_{k=0}^{L-2} \frac{\exp[\psi\{\alpha_{ij}^{-1}(Q_{ij}(t_{Kk+m}) - \theta)\}]}{\exp\{\alpha_{ij}^{-1} B_{ij}(t_{Kk+m})\}} \times E \left[\frac{\exp[\psi\{\alpha_{ij}^{-1}(Q_{ij}(t_{K(L-1)+m}) - \theta)\}]}{\exp\{\alpha_{ij}^{-1} B_{ij}(t_{K(L-1)+m})\}} \middle| \mathcal{F}_{t_{K(L-1)+m}} \right] \right]^{1/K} \\ &\leq \prod_{m=0}^{K-1} \mathbb{E} \left[\prod_{k=0}^{L-2} \frac{\exp[\psi\{\alpha_{ij}^{-1}(Q_{ij}(t_{Kk+m}) - \theta)\}]}{\exp\{\alpha_{ij}^{-1} B_{ij}(t_{Kk+m})\}} \right]^{1/K} \delta^{1/(KL)} \\ &\leq \delta, \end{aligned} \tag{7.3}$$

where the third and fourth inequalities are due to the Hölder's inequality and (7.4), respectively, and the last inequality can be derived recursively. Since $-\log(1-x+x^2) \leq \log(1+x+x^2)$, we have

$$\begin{aligned}
& \mathbb{E} \left[\frac{\exp [\psi \{ \alpha_{ij}^{-1} (Q_{ij}(t_k) - \theta) \}]}{\exp \{ \alpha_{ij}^{-1} B_{ij}(t_k) \}} \middle| \mathcal{F}_{t_k} \right] \leq \mathbb{E} \left[\frac{1 + \alpha_{ij}^{-1} (Q_{ij}(t_k) - \theta) + \alpha_{ij}^{-2} (Q_{ij}(t_k) - \theta)^2}{\exp \{ \alpha_{ij}^{-1} B_{ij}(t_k) \}} \middle| \mathcal{F}_{t_k} \right] \quad (7.4) \\
& = \frac{1 + \alpha_{ij}^{-1} (A_{ij}(t_k) - \theta) + \alpha_{ij}^{-2} \{ V_{ij}(t_k) + (A_{ij}(t_k) - \theta)^2 \}}{\exp \{ \alpha_{ij}^{-1} B_{ij}(t_k) \}} \\
& \leq \frac{\exp [\alpha_{ij}^{-1} (A_{ij}(t_k) - \theta) + \alpha_{ij}^{-2} \{ V_{ij}(t_k) + (A_{ij}(t_k) - \theta)^2 \}]}{\exp \{ \alpha_{ij}^{-1} B_{ij}(t_k) \}} \leq \delta^{1/L} \text{ a.s.},
\end{aligned}$$

where the last equality is due to (7.1) and the definition of $V_{ij}(t_k)$. Similarly to the proof of (7.3), we can show

$$\Pr \{ f(\theta) < B_-(\theta) \} \leq \delta. \quad (7.5)$$

From (7.2) and the fact that $|A_{ij}(t_k)| \leq \nu_\gamma + \nu_e \zeta n / (\phi K)$, we have for large n ,

$$\alpha_{ij}^{-2} V_{ij}^* + \alpha_{ij}^{-2} \{ A_{ij}^{*(2)} - (A_{ij}^*)^2 \} + \frac{\log \delta^{-1}}{L} < \frac{1}{4}. \quad (7.6)$$

Let θ_+ be the smallest solution of the quadratic equation $B_+(\theta) = 0$ and θ_- be the largest solution of the quadratic equation $B_-(\theta) = 0$. First calculate θ_+ . We have

$$B_+(\theta) = \alpha_{ij}^{-1} \theta^2 - \theta (1 + 2\alpha_{ij}^{-1} A_{ij}^*) + A_{ij}^* + \alpha_{ij}^{-1} V_{ij}^* + \alpha_{ij}^{-1} A_{ij}^{*(2)} + \frac{\log(1/\delta)}{L \alpha_{ij}^{-1}},$$

where

$$A_{ij}^* = \frac{1}{n-K+1} \sum_{k=0}^{n-K} A_{ij}(t_k) = \frac{1}{\phi K} \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) \right\}^2 \sum_{k=0}^{n-K} \mathbb{E} \left[\int_{t_{k+l}}^{t_{k+l+1}} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right] + \frac{n-K+1}{\phi K} \zeta \eta_{ij},$$

$$V_{ij}^* = \frac{1}{n-K+1} \sum_{k=0}^{n-K} V_{ij}(t_k), \quad \text{and} \quad A_{ij}^{*(2)} = \frac{1}{n-K+1} \sum_{k=0}^{n-K} A_{ij}^2(t_k).$$

Then we have

$$\begin{aligned}
\theta_+ &= A_{ij}^* + 2 \left(\alpha_{ij}^{-1} V_{ij}^* + \alpha_{ij}^{-1} \{A_{ij}^*(2) - (A_{ij}^*)^2\} + \frac{\log \delta^{-1}}{L \alpha_{ij}^{-1}} \right) \\
&\times \left(1 + \sqrt{1 - 4 \left(\alpha_{ij}^{-2} V_{ij}^* + \alpha_{ij}^{-2} \{A_{ij}^*(2) - (A_{ij}^*)^2\} + \frac{\log \delta^{-1}}{L} \right)} \right)^{-1} \\
&\leq A_{ij}^* + 2 \left(\alpha_{ij}^{-1} V_{ij}^* + \alpha_{ij}^{-1} \{A_{ij}^*(2) - (A_{ij}^*)^2\} + \frac{\log \delta^{-1}}{L \alpha_{ij}^{-1}} \right).
\end{aligned}$$

Similarly, we have

$$\theta_- \geq A_{ij}^* - 2 \left(\alpha_{ij}^{-1} V_{ij}^* + \alpha_{ij}^{-1} \{A_{ij}^*(2) - (A_{ij}^*)^2\} + \frac{\log \delta^{-1}}{L \alpha_{ij}^{-1}} \right).$$

Since $\theta_{\pm}$ are random variables, we cannot use above results directly. Now, we handle the randomness of $\theta_{\pm}$. For large n , we have

$$-2 \left(\nu_{\gamma} + \frac{n-K+1}{\phi K} \zeta \nu_{\varepsilon} \right) < \theta_- < \theta_+ < 2 \left(\nu_{\gamma} + \frac{n-K+1}{\phi K} \zeta \nu_{\varepsilon} \right) \text{ a.s.}$$

Without loss of the generality, we assume that $s = \sqrt{L}$ is an integer value. Define

$$\mathcal{E} = \{x_0, \dots, x_s\},$$

where

$$x_0 = -2 \left(\nu_{\gamma} + \frac{n-K+1}{\phi K} \zeta \nu_{\varepsilon} \right), \quad x_s = 2 \left(\nu_{\gamma} + \frac{n-K+1}{\phi K} \zeta \nu_{\varepsilon} \right)$$

and

$$x_{i+1} - x_i = \frac{4}{s} \left(\nu_{\gamma} + \frac{n-K+1}{\phi K} \zeta \nu_{\varepsilon} \right) \text{ for } i = 0, \dots, s-1.$$

Let

$$E_{1,i} = \left\{ \theta_- - \frac{4}{s} \left(\nu_{\gamma} + \frac{n-K+1}{\phi K} \zeta \nu_{\varepsilon} \right) < x_i \leq \theta_- \right\}, \quad E_{2,i} = \{B_-(x_i) \leq f(x_i)\}$$

$$E_{2,i} = \left\{ \theta_+ \leq x_i < \theta_+ + \frac{4}{s} \left(\nu_{\gamma} + \frac{n-K+1}{\phi K} \zeta \nu_{\varepsilon} \right) \right\}, \quad E_{4,i} = \{f(x_i) \leq B_+(x_i)\}.$$

We have

$$\Pr\{\cup_{i=0}^s \cup_{j=0}^s (E_{1,i} \cap E_{2,i} \cap E_{3,j} \cap E_{4,j})\} = \Pr[\{\cup_{i=0}^s (E_{1,i} \cap E_{2,i})\} \cap \{\cup_{j=0}^s (E_{3,j} \cap E_{4,j})\}] \geq 1 - \Pr[\{\cup_{i=0}^s (E_{1,i} \cap E_{2,i})\}^c] - \Pr[\{\cup_{j=0}^s (E_{3,j} \cap E_{4,j})\}^c] \geq 1 - 2(s+1)\delta,$$

$$(7.7)$$

where the last inequality is due to (7.8) and (7.9) below. Since $E_{1,i}$, $i = 0, \dots, s$, are mutually exclusive, we have

$$\begin{aligned} \Pr\{\cup_{i=0}^s (E_{1,i} \cap E_{2,i})\} &= \sum_{i=0}^s \Pr(E_{1,i} \cap E_{2,i}) \quad (7.8) \\ &\geq (s+1) - \sum_{i=0}^s \{\Pr(E_{1,i}^c) + \Pr(E_{2,i}^c)\} \\ &\geq (s+1) - (s+1)\delta - \sum_{i=0}^s \Pr(E_{1,i}^c) \\ &\geq (s+1) - (s+1)\delta - \left\{s+1 - \sum_{i=0}^s \Pr(E_{1,i})\right\} \\ &\geq 1 - (s+1)\delta, \end{aligned}$$

where the second inequality is due to (7.5). Similarly,

$$\Pr\{\cup_{i=0}^s (E_{3,i} \cap E_{4,i})\} \geq 1 - (s+1)\delta. \quad (7.9)$$

Thus, with probability greater than $1 - 2(s+1)\delta$, there exist $x_L, x_U \in \mathcal{G}$ such that

$$\theta_- - \frac{4}{s} \left(\nu_\gamma + \frac{n-K+1}{\phi K} \zeta \nu_\epsilon \right) < x_L \leq \theta_-, \quad B_-(x_L) \leq f(x_L)$$

$$\theta_+ \leq x_U < \theta_+ + \frac{4}{s} \left(\nu_\gamma + \frac{n-K+1}{\phi K} \zeta \nu_\epsilon \right), \quad f(x_U) \leq B_+(x_U).$$

For large n , we can easily show that

$$B_+(x_U) \leq 0 \quad \text{and} \quad 0 \leq B_-(x_L) \quad \text{a.s.}$$

Therefore, since $f(\theta)$ is non-increasing function with respect to θ , with probability greater than $1 - 2(s+1)\delta$

$$\theta_- - \frac{4}{s} \left(\nu_\gamma + \frac{n-K+1}{\phi K} \zeta \nu_\varepsilon \right) < x_L \leq \hat{\theta}_{ij, \alpha} \leq x_U < \theta_+ + \frac{4}{s} \left(\nu_\gamma + \frac{n-K+1}{\phi K} \zeta \nu_\varepsilon \right),$$

which implies

$$|\hat{\theta}_{ij, \alpha} - A_{ij}^*| \leq 2 \left(\alpha_{ij}^{-1} V_{ij}^* + \alpha_{ij}^{-1} \{A_{ij}^{*(2)} - (A_{ij}^*)^2\} + \frac{\log \delta^{-1}}{L \alpha_{ij}^{-1}} \right) + \frac{8}{s} \left(\nu_\gamma + \frac{n-K+1}{\phi K} \zeta \nu_\varepsilon \right). \quad (7.10)$$

Now we investigate A_{ij}^* . Since $\sum_{k=0}^{n-K} \left[E \left\{ \int_{t_{k+l}}^{t_{k+l+1}} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right\} - \int_{t_{k+l}}^{t_{k+l+1}} \gamma_{ij}(t) dt \right]$ is a martingale and $\gamma_{ij}(t)$ is bounded, we can show

$$\Pr \left(\left| \sum_{k=0}^{n-K} \left[E \left\{ \int_{t_{k+l}}^{t_{k+l+1}} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right\} - \int_{t_{k+l}}^{t_{k+l+1}} \gamma_{ij}(t) dt \right] \right| \geq C \sqrt{\frac{\log \delta^{-1}}{n}} \right) \leq \delta.$$

Also, simple algebraic manipulations show

$$|A_{ij}^* - \theta_{ij,0}| \leq \left| \frac{1}{\phi K} \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) \right\}^{2n-K} \sum_{k=0}^{n-K} \left[E \left\{ \int_{t_{k+l}}^{t_{k+l+1}} \gamma_{ij}(t) dt \middle| \mathcal{F}_{t_k} \right\} - \int_{t_{k+l}}^{t_{k+l+1}} \gamma_{ij}(t) dt \right] \right| + 2 \frac{K}{n} \nu_\gamma.$$

These imply that, with probability at least $1 - \delta$,

$$|A_{ij}^* - \theta_{ij,0}| \leq C \sqrt{\frac{\log \delta^{-1}}{n}} + 2 \frac{K}{n} \nu_\gamma. \quad (7.11)$$

Collecting (7.7), (7.10), and (7.11), we have with probability greater than $1 - 2(s+2)\delta$,

$$|\hat{\theta}_{ij, \alpha} - \theta_{ij,0}| \leq C \sqrt{\frac{(\nu_\gamma + \frac{n\zeta}{\phi K} V_\varepsilon)^4 K \log \delta^{-1}}{(n-K+1) \bar{V}_{ij}^*}} + 2 \nu_\gamma \frac{K}{n}. \quad (7.12)$$

$\hat{\eta}_{ii}$ in (3.5) is the same as the solution of the following equation,

$$\sum_{k=0}^{n-1} \psi(\alpha_{\eta, ii} [\{Y_i(t_{k+1}) - Y_i(t_k)\}^2/2 - \theta]) = 0. \quad (7.13)$$

So we use $\hat{\eta}_{ii}$ is as the solution of (7.13).

Proof of Theorem 3.2—Define

$$f_{\eta}(\theta) = \frac{1}{\alpha_{\eta, ii}^n} \sum_{k=0}^{n-1} \psi[\alpha_{\eta, ii} \{Q_{\eta, ii}(t_k) - \theta\}],$$

$$A_{\eta, ii}(t_k) = \frac{1}{2} \left[E \left\{ \int_{t_k}^{t_k+1} \gamma_{ii}(t) dt \middle| \mathcal{F}_{t_k} \right\} + 2\eta_{ii} \right],$$

$$V_{\eta, ii}(t_k) = E[\{Q_{\eta, ii}(t_k) - A_{\eta, ii}(t_k)\}^2 | \mathcal{F}_{t_k}],$$

where

$$Q_{\eta, ii}(t_k) = \frac{1}{2} \{Y_{t(t_k+1)} - Y_{t(t_k)}\}^2.$$

Then

$$A_{\eta, ii}^* = \frac{1}{n} \sum_{k=0}^{n-1} A_{\eta, ii}(t_k) = \eta_{ii} + \frac{1}{2n} \sum_{k=0}^{n-1} E \left[\int_{t_k}^{t_k+1} \gamma_{ii}(t) dt \middle| \mathcal{F}_{t_k} \right] \text{ a.s.},$$

$$V_{\eta, ii}^* = \frac{1}{n} \sum_{k=0}^{n-1} V_{\eta, ii}(t_k) \leq 4\nu_e^2 + Cn^{-1} \text{ a.s.}$$

Similar to the proofs of Theorem 3.1, we can show that for large n ,

$$\Pr \left[|\hat{\eta}_{ii} - \eta_{ii}^*| \geq C \sqrt{\frac{\nu_e^4 \log \delta^{-1}}{n \bar{V}_{ij, \epsilon}^*}} + \frac{1}{2n} \nu_e \right] \leq 2\sqrt{n}\delta.$$

Then (3.6) is proved by Theorem 3.1.

7.2 Proofs of Section 4

Proposition 7.1—The concentration inequality (4.2), sparse condition (4.1) and Assumption 2(a) are satisfied. Then with probability greater than $1 - p^{-1}$, we have

$$\max_{1 \leq i \leq r} |\hat{\lambda}_i / \lambda_i - 1| \leq C \frac{p}{\lambda_r} \sqrt{\frac{\log(p \vee n)}{n^{1/2}}}, \quad (7.14)$$

$$\|\hat{\mathbf{q}}_i - \text{sign}(\langle \hat{\mathbf{q}}_i, \bar{\mathbf{q}}_i \rangle) \bar{\mathbf{q}}_i\|_\infty \leq \frac{C}{D_\lambda \sqrt{p}} \left(M_{\sigma^2 p} + p \sqrt{\frac{\log(p \vee n)}{n^{1/2}}} \right). \quad (7.15)$$

Proof of Proposition 7.1—First, consider (7.14). By Weyl's theorem, we have with probability greater than $1 - p^{-1}$,

$$\max_{1 \leq j \leq p} |\hat{\lambda}_j - \lambda_j| \leq \|\Gamma - \hat{\Gamma}\|_2 \leq \|\Gamma - \hat{\Gamma}\|_1 \leq p \|\Gamma - \hat{\Gamma}\|_{\max} \leq Cp \sqrt{\frac{\log(p \vee n)}{n^{1/2}}}.$$

Thus, with probability greater than $1 - p^{-1}$,

$$\max_{1 \leq j \leq r} |\hat{\lambda}_j / \lambda_j - 1| \leq C \frac{p}{\lambda_r} \sqrt{\frac{\log(p \vee n)}{n^{1/2}}}.$$

Consider (7.15). By Theorem 2.1 (Fan et al., 2016c), we have with probability greater than $1 - p^{-1}$,

$$\|\hat{\mathbf{q}}_i - \text{sign}(\langle \hat{\mathbf{q}}_i, \bar{\mathbf{q}}_i \rangle) \bar{\mathbf{q}}_i\|_\infty \leq C \frac{\|\Sigma\|_1 + \|\hat{\Gamma} - \Gamma\|_1}{D_\lambda \sqrt{p}} \leq \frac{C}{D_\lambda \sqrt{p}} \left(M_{\sigma^2 p} + p \sqrt{\frac{\log(p \vee n)}{n^{1/2}}} \right),$$

where the last inequality is due to (4.2) and (7.16) below. We have

$$\|\Sigma\|_1 = \max_{1 \leq i \leq p} \sum_{j=1}^p |\sum_{ij}| \leq \max_{1 \leq i \leq p} \sum_{j=1}^p |\sum_{ii} \sum_{jj}|^{(1-q)/2} |\sum_{ij}|^q \leq M_{\sigma^2 p}. \quad (7.16)$$

Proof of Theorem 4.1—First, consider (4.4) and (4.6). We have, with probability greater than $1 - p^{-1}$,

$$\begin{aligned} & \left\| \sum_{k=1}^r \hat{\lambda}_k \hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^\top - \sum_{k=1}^r \bar{\lambda}_k \bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^\top \right\|_{\max} \leq \left\| \sum_{k=1}^r \hat{\lambda}_k (\hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^\top - \bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^\top) - \sum_{k=1}^r (\bar{\lambda}_k - \hat{\lambda}_k) \bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^\top \right\|_{\max} \quad (7.17) \\ & \leq \sum_{k=1}^r |(\hat{\lambda}_k - \lambda_k) + \lambda_k| \|\hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^\top - \bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^\top\|_{\max} + \max_{1 \leq i \leq r} |\bar{\lambda}_i - \hat{\lambda}_i| \sum_{k=1}^r \|\bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^\top\|_{\max} \\ & \leq C \sum_{k=1}^r |(\hat{\lambda}_k - \lambda_k) + \lambda_k| \left\{ p^{-1/2} \max_{1 \leq j \leq r} \|\hat{\mathbf{q}}_j - \text{sign}(\langle \bar{\mathbf{q}}_j, \hat{\mathbf{q}}_j \rangle) \bar{\mathbf{q}}_j\|_\infty \right. \\ & \quad \left. + \max_{1 \leq j \leq r} \|\hat{\mathbf{q}}_j - \text{sign}(\langle \bar{\mathbf{q}}_j, \hat{\mathbf{q}}_{kj} \rangle) \bar{\mathbf{q}}_j\|_\infty^2 \right\} + Cp^{-1} (\|\Gamma - \hat{\Gamma}\|_1 + \|\Sigma\|_1) \\ & \leq C \left(M_{\sigma^2 p} + \sqrt{\frac{\log(p \vee n)}{n^{1/2}}} \right), \end{aligned}$$

where the third inequality is from the Weyl's theorem and Assumption 2(a), and the last inequality is due to Proposition 7.1. Simple algebra shows that, with probability greater than $1 - p^{-1}$,

$$\begin{aligned} \|\tilde{\Sigma} - \Sigma\|_{\max} &= \|\hat{\Gamma} - \Gamma - \sum_{k=1}^r \hat{\lambda}_k \hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^\top + \sum_{k=1}^r \bar{\lambda}_k \bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^\top\| \\ &\leq \|\hat{\Gamma} - \Gamma\|_{\max} + \left\| \sum_{k=1}^r \hat{\lambda}_k \hat{\mathbf{q}}_k \hat{\mathbf{q}}_k^\top - \sum_{k=1}^r \bar{\lambda}_k \bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^\top \right\|_{\max} \leq C \left(\frac{M_{\sigma} s_p}{p} + \sqrt{\frac{\log(p \vee n)}{n^{1/2}}} \right), \end{aligned} \quad (7.18)$$

where the third line is from (7.17) and Proposition 7.1. From (7.17) and (7.18), statements (4.4) and (4.6) can be shown.

Consider (4.3). We have, with probability $1 - 2p^{-1}$,

$$\begin{aligned} &\max_{1 \leq i, j \leq p} \frac{|\sum_{ii} \tilde{\Sigma} \vee 0)(\sum_{jj} \tilde{\Sigma} \vee 0) - \sum_{ii} \sum_{jj}|}{\sum_{ii} \sum_{jj}} \\ &\leq \max_{1 \leq i, j \leq p} \frac{|\sum_{ii} \tilde{\Sigma} - \sum_{ii}| |\sum_{jj} \tilde{\Sigma} - \sum_{jj}| + |\sum_{ii} \tilde{\Sigma} - \sum_{ii}| \sum_{jj} + |\sum_{jj} \tilde{\Sigma} - \sum_{jj}| \sum_{ii}}{\sum_{ii} \sum_{jj}} \leq C(\beta_n^2 + \beta_n) \\ &\leq C \left(\frac{s_p}{\sqrt{p}} + \sqrt{\frac{\log(p \vee n)}{n^{1/2}}} \right) = o(1), \end{aligned}$$

where the second inequality is due to (7.18), and the last inequality is due to the fact that by Chebyshev's inequality, we have

$$\Pr \left(M_{\sigma} \geq \sqrt{p \mathbb{E}(M_{\sigma}^2)} \right) \leq \frac{\mathbb{E}(M_{\sigma}^2)}{\mathbb{E}(M_{\sigma}^2)p} = p^{-1}.$$

Then by taking large C_{ω} , we have for large n ,

$$\Pr \left(\|\tilde{\Sigma} - \Sigma\|_{\max} \geq \omega_{ij}/2 \right) \leq \Pr \left(\|\tilde{\Sigma} - \Sigma\|_{\max} > CC_{\omega} \beta_n \right) + \Pr \left(M_{\sigma} \geq \sqrt{p \mathbb{E}(M_{\sigma}^2)} \right) \leq 2p^{-1}.$$

On the event

$$\begin{aligned} E &= \{|\sum_{ij} \tilde{\Sigma} - \sum_{ij}| < \omega_{ij}/2 \text{ for all } i, j = 1, \dots, p\} \cap \{|\sum_{ii} \tilde{\Sigma} \vee 0)(\sum_{jj} \tilde{\Sigma} \vee 0) - \sum_{ii} \sum_{jj}| / \sum_{ii} \sum_{jj}, \\ &\leq C \left(s_p / \sqrt{p} + \sqrt{\log(p \vee n) / n^{1/2}} \right) \text{ for all } i, j = 1, \dots, p\} \end{aligned}$$

we have

$$\begin{aligned}
\sum_{i=1}^p |\widehat{\Sigma}_{ij} - \Sigma_{ij}| &= \sum_{i=1}^p |s_{ij}(\widetilde{\Sigma}_{ij}) \mathbf{1}(|\widetilde{\Sigma}_{ij}| \geq \varpi_{ij}) - \Sigma_{ij}| \\
&\leq \sum_{i=1}^p |s_{ij}(\widetilde{\Sigma}_{ij}) - \Sigma_{ij}| \mathbf{1}(|\widetilde{\Sigma}_{ij}| \geq \varpi_{ij}) + \sum_{i=1}^p |\Sigma_{ij}| \mathbf{1}(|\Sigma_{ij}| < \varpi_{ij}) \\
&+ \sum_{i=1}^p |\Sigma_{ij}| \mathbf{1}(|\widetilde{\Sigma}_{ij}| \geq \varpi_{ij}) - \mathbf{1}(|\Sigma_{ij}| \geq \varpi_{ij}) \leq \frac{3}{2} \sum_{i=1}^p \varpi_{ij} \mathbf{1}(|\Sigma_{ij}| \geq \varpi_{ij} - |\Sigma_{ij} - \widetilde{\Sigma}_{ij}|) \\
&+ \sum_{i=1}^p |\Sigma_{ij}|^q \varpi_{ij}^{1-q} + \sum_{i=1}^p |\Sigma_{ij}| \mathbf{1}\left(\left||\widetilde{\Sigma}_{ij}| - \varpi_{ij}\right| \leq |\Sigma_{ij} - \widetilde{\Sigma}_{ij}|\right) \\
&\leq \frac{3}{2} 2^q \sum_{i=1}^p |\Sigma_{ij}|^q \varpi_{ij}^{1-q} + \sum_{i=1}^p |\Sigma_{ij}|^q \varpi_{ij}^{1-q} + \sum_{i=1}^p |\Sigma_{ij}| \mathbf{1}\left(|\Sigma_{ij}| \leq \frac{3}{2} \varpi_{ij}\right) \\
&\leq C \sum_{i=1}^p |\Sigma_{ij}|^q \varpi_{ij}^{1-q} \leq CM_{\sigma^s p} \beta_n^{1-q}.
\end{aligned}$$

Then (4.3) is proved.

Consider (4.6). We have

$$\|\widehat{\Gamma}_{POET} - \Gamma\|_{\Gamma} \leq \left\| \sum_{k=1}^r \widehat{\lambda}_k \widehat{\mathbf{q}}_k \widehat{\mathbf{q}}_k^{\top} - \Theta \right\|_{\Gamma} + \|\widehat{\Sigma} - \Sigma\|_{\Gamma} = I + II.$$

For I , similar to the proof of Theorem 4.2 in Kim et al. (2015b), we can show, with probability greater than $1 - p^{-1}$,

$$\begin{aligned}
I &\leq C \left(p^{-1/2} \max_{1 \leq k \leq r} \frac{|\widehat{\lambda}_k - \bar{\lambda}_k|}{\bar{\lambda}_k} + p^{1/2} \widehat{\lambda}_1 \max_{1 \leq k \leq r} \|\widehat{\mathbf{q}}_k - \text{sign}(\langle \widehat{\mathbf{q}}_k, \widehat{\mathbf{q}}_k \rangle) \mathbf{q}_k\|_{\max}^2 \right. \\
&\quad \left. + \max_{1 \leq j \leq r} \frac{\widehat{\lambda}_j}{\bar{\lambda}_j^{1/2}} \max_{1 \leq k \leq r} \|\widehat{\mathbf{q}}_k - \text{sign}(\langle \widehat{\mathbf{q}}_k, \widehat{\mathbf{q}}_k \rangle) \mathbf{q}_k\|_{\max} \right) \leq C \left(\frac{p^{1/2}}{\bar{\lambda}_r} \beta_n + \frac{p^{3/2} \lambda_1}{D_{\lambda}^2} \beta_n^2 + \frac{\lambda_1}{D_{\lambda}} \beta_n \right) \\
&\leq C \left(\beta_n + \frac{p^{3/2}}{D_{\lambda}} \beta_n^2 \right),
\end{aligned}$$

where the second inequality is due to Proposition 7.1 and Weyl's theorem. For II , by (4.3), we have, with probability greater than $1 - 2p^{-1}$

$$II^2 = p^{-1} \|(\widehat{\Sigma} - \Sigma) \Gamma^{-1}\|_F^2 \leq p^{-1} \|\widehat{\Sigma} - \Sigma\|_2^2 \|\Gamma^{-1}\|_F^2 \leq CM_{\sigma^s p}^2 \beta_n^{2-2q}.$$

Therefore, we have, with probability greater than $1 - 2p^{-1}$

$$\|\widehat{\Gamma}_{POET} - \Gamma\|_{\Gamma} \leq C \left(\beta_n + \frac{p^{3/2}}{D_{\lambda}} \beta_n^2 + M_{\sigma^s p} \beta_n^{1-q} \right) \leq C \left(\frac{p^{1/2} \log(p \vee n)}{n^{1/2}} + M_{\sigma^s p} \beta_n^{1-q} \right).$$

Finally, consider (4.7) and (4.8). For (4.7), since the minimum eigenvalue of Σ stay away from 0, the minimum eigenvalue of $\hat{\Sigma}$ is bounded away from 0 with probability approaching 1. Thus, $\hat{\Sigma}^{-1}$ exists and we have

$$\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_2 \leq CM_{\sigma^2 p} \beta_n^{1-q}.$$

For (4.8), let

$$\Theta = \bar{Q}\Xi\bar{Q}^\top \text{ and } \hat{\Theta} = \hat{Q}\hat{\Xi}\hat{Q}^\top,$$

where $\Xi = \text{diag}(\bar{\lambda}_1, \dots, \bar{\lambda}_r)$, $\hat{\Xi} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_r)$, $\bar{Q} = (\bar{\mathbf{q}}_1, \dots, \bar{\mathbf{q}}_r)$, and $\hat{Q} = (\hat{\mathbf{q}}_1, \dots, \hat{\mathbf{q}}_r)$. Define

$$\mathbf{H} = \bar{Q}\Xi^{1/2}, \quad \hat{\mathbf{H}} = \hat{Q}\hat{\Xi}^{1/2},$$

and

$$\Gamma_1 = \hat{\mathbf{H}}\hat{\mathbf{H}}^\top + \Sigma.$$

Then we have

$$\|\hat{\Gamma}^{-1} - \Gamma^{-1}\|_2 \leq \|\hat{\Gamma}^{-1} - \Gamma_1^{-1}\|_2 + \|\Gamma_1^{-1} - \Gamma^{-1}\|_2.$$

First consider $\|\hat{\Gamma}^{-1} - \Gamma_1^{-1}\|_2$. We have

$$\|\hat{\mathbf{H}}\|_2 \leq \hat{\lambda}_1^{1/2}, \quad (7.19)$$

and since

$$\lambda_{\min}(\hat{\mathbf{H}}^\top \hat{\Sigma}^{-1} \hat{\mathbf{H}}) \geq \lambda_{\min}(\hat{\Sigma}^{-1}) \hat{\lambda}_r,$$

we have

$$\|\mathbf{G}\|_2 \leq \frac{1}{1 + \lambda_{\min}(\hat{\Sigma}^{-1}) \hat{\lambda}_r}, \quad (7.20)$$

where $\mathbf{G} = (\mathbf{I}_r + \hat{\mathbf{H}}^\top \hat{\Sigma}^{-1} \hat{\mathbf{H}})^{-1}$, and $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ are the smallest and largest eigenvalues of $\mathbf{A}$, respectively. Similarly,

$$\|\mathbf{G}_1\|_2 \leq \frac{1}{1 + \lambda_{\min}(\mathbf{\Sigma}^{-1})\hat{\lambda}_r}. \quad (7.21)$$

where $\mathbf{G}_1 = (\mathbf{I}_r + \hat{\mathbf{H}}^\top \mathbf{\Sigma}^{-1} \hat{\mathbf{H}})^{-1}$. Then, by Sherman-Morrison-Woodbury formula, we have with probability approaching 1

$$\begin{aligned} \|\hat{\mathbf{\Gamma}}^{-1} - \mathbf{\Gamma}_1^{-1}\|_2 &\leq \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 + \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 \|\hat{\mathbf{H}}(\mathbf{I}_r + \hat{\mathbf{H}}^\top \widehat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^\top \widehat{\mathbf{\Sigma}}^{-1}\|_2 \\ &\quad + \|\mathbf{\Sigma}^{-1} \hat{\mathbf{H}}(\mathbf{I}_r + \hat{\mathbf{H}}^\top \widehat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^\top\|_2 \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 \\ &\quad + \|\mathbf{\Sigma}^{-1} \hat{\mathbf{H}}\{(\mathbf{I}_r + \hat{\mathbf{H}}^\top \widehat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}})^{-1} - (\mathbf{I}_r + \hat{\mathbf{H}} \mathbf{\Sigma}^{-1} \hat{\mathbf{H}}^\top)^{-1}\} \hat{\mathbf{H}}^\top \mathbf{\Sigma}^{-1}\|_2 \\ &\leq \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 \{1 + \|\hat{\mathbf{H}}\|_2^2 \|\mathbf{I}_r + \hat{\mathbf{H}}^\top \widehat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{H}}\|_2 (\|\widehat{\mathbf{\Sigma}}^{-1}\|_2 + \|\mathbf{\Sigma}^{-1}\|_2)\} \\ &\quad + \|\mathbf{\Sigma}^{-1}\|_2^2 \|\hat{\mathbf{H}}\|_2^2 \|\mathbf{G} - \mathbf{G}_1\|_2 \leq C \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 \left[1 + \frac{\hat{\lambda}_1 \{\lambda_{\max}(\mathbf{\Sigma}^{\wedge-1}) + \lambda_{\max}(\mathbf{\Sigma}^{-1})\}}{\lambda_{\min}(\mathbf{\Sigma}^{\wedge-1}) \hat{\lambda}_r} \right] \\ &\quad + C \frac{\hat{\lambda}_1}{\lambda_{\min}(\mathbf{\Sigma})^2} \|\mathbf{G} - \mathbf{G}_1\|_2 \leq C \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 + C \frac{\hat{\lambda}_1}{\lambda_{\min}(\mathbf{\Sigma})^2} \|\mathbf{G}\|_2 \|\mathbf{G}_1\|_2 \|\mathbf{G}^{-1} - \mathbf{G}_1^{-1}\|_2 \\ &\leq C \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 \left[1 + \frac{\hat{\lambda}_1^2 \lambda_{\max}(\mathbf{\Sigma})^2}{\lambda_{\min}(\mathbf{\Sigma})^2 \hat{\lambda}_r^2} \right] \leq C \|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2, \end{aligned} \quad (7.22)$$

where the third and fifth inequalities are due to (7.19)–(7.21), and the fourth and last inequalities are from the Weyl's theorem, Proposition 7.1, and (4.3).

Consider $\|\mathbf{\Gamma}_1^{-1} - \mathbf{\Gamma}^{-1}\|_2$. We have with probability approaching 1,

$$\|\mathbf{H} - \hat{\mathbf{H}}\|_2 \leq \|\hat{\mathbf{Q}} - \bar{\mathbf{Q}}\|_2 \lambda_1^{1/2} + \max_{1 \leq k \leq r} |\hat{\lambda}_k^{1/2} - \bar{\lambda}_k^{1/2}| \leq C \left(\frac{\lambda_1^{1/2} p}{D_\lambda} \beta_n + \frac{p}{\lambda_1^{1/2}} \beta_n \right),$$

where the last inequality is due to the Weyl's theorem and Proposition 7.1. Then, similar to the proofs of (7.22), we can show that with probability approaching 1,

$$\begin{aligned}
& \|\Gamma_1^{-1} - \Gamma^{-1}\|_2 \leq \|\Sigma^{-1}\|_2^2 \|\hat{\mathbf{H}} - \mathbf{H}\|_2 \|\mathbf{I}_r + \hat{\mathbf{H}}^\top \Sigma^{-1} \hat{\mathbf{H}}\|_2^{-1} \|\hat{\mathbf{H}}\|_2 \\
& + \|\Sigma^{-1}\|_2^2 \|\mathbf{H}\|_2^2 \|\mathbf{I}_r + \hat{\mathbf{H}}^\top \Sigma^{-1} \hat{\mathbf{H}}\|_2^{-1} - (\mathbf{I}_r + \mathbf{H}^\top \Sigma^{-1} \mathbf{H})^{-1}\|_2 \\
& + \|\Sigma^{-1}\|_2^2 \|\mathbf{H}\|_2 \|\mathbf{I}_r + \hat{\mathbf{H}}^\top \Sigma^{-1} \hat{\mathbf{H}}\|_2^{-1} \|\hat{\mathbf{H}} - \mathbf{H}\|_2 \leq C\beta_n + \frac{C}{\lambda_r} \|\hat{\mathbf{H}}^\top \Sigma^{-1} \hat{\mathbf{H}} - \mathbf{H}^\top \Sigma^{-1} \mathbf{H}\|_2 \leq C\beta_n.
\end{aligned}$$

References

- Ahn SC, Horenstein AR. 2013; Eigenvalue ratio test for the number of factors. *Econometrica*. 81(3): 1203–1227.
- Aït-Sahalia Y, Fan J, Xiu D. 2010; High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*. 105(492):1504–1517.
- Ait-Sahalia Y, Xiu D. 2015 Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Chicago Booth Research Paper*. :15–43.
- Alessi L, Barigozzi M, Capasso M. 2010; Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*. 80(23):1806–1813.
- Bai J, Ng S. 2002; Determining the number of factors in approximate factor models. *Econometrica*. 70(1):191–221.
- Barndorff-Nielsen OE. 2002; Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 64(2):253–280.
- Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2008; Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*. 76(6):1481–1536.
- Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2011; Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*. 162(2):149–169.
- Bibinger M, Hautsch N, Malec P, Reiß M, et al. 2014; Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *The Annals of Statistics*. 42(4): 1312–1346.
- Candès EJ, Li X, Ma Y, Wright J. 2011; Robust principal component analysis? *Journal of the ACM (JACM)*. 58(3):11.
- Catoni O. 2012; Challenging the empirical mean and empirical variance: a deviation study. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. 48(4):1148–1185.
- Christensen K, Kinnebrock S, Podolskij M. 2010; Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*. 159(1): 116–133.
- Cox JC, Ingersoll JE Jr, Ross SA. 1985; A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*. 53:385–407.
- Doz C, Giannone D, Reichlin L. 2011; A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*. 164(1):188–205.
- Fan J, Furger A, Xiu D. 2016a; Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. *Journal of Business & Economic Statistics*. 34:489–503.
- Fan J, Li Q, Wang Y. 2016b Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
- Fan J, Liao Y, Mincheva M. 2013; Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 75(4): 603–680.

- Fan J, Wang W, Zhong Y. 2016c An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. arXiv preprint arXiv:1603.03516.
- Fan J, Wang Y. 2007; Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*. 102(480):1349–1362.
- Hallin M, Liška R. 2007; Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*. 102(478):603–617.
- Huang X, Tauchen G. 2005; The relative contribution of jumps to total price variance. *Journal of financial econometrics*. 3(4):456–499.
- Huber PJ. 1964; Robust estimation of a location parameter. *The Annals of Mathematical Statistics*. 35(1):73–101.
- Jacod J, Li Y, Mykland PA, Podolskij M, Vetter M. 2009; Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*. 119(7):2249–2276.
- Jacod J, Protter P. 2012 Discretization of processes.
- Kim D, Kong X-B, Li C-X, Wang Y. 2015a Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data.
- Kim D, Liu Y, Wang Y. 2015b Large volatility matrix estimation with factor-based diffusion model for high-frequency financial data.
- Kim D, Wang Y. 2016; Sparse pca based on high-dimensional $it\hat{o}$ processes with measurement errors. *Journal of Multivariate Analysis*. 152:172–18.
- Lam C, Yao Q. 2012 Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*. :694–726.
- Minsker S. 2016 Sub-gaussian estimators of the mean of a random matrix with heavy-tailed entries. arXiv preprint arXiv:1605.07129.
- Mykland PA, Zhang L. 2016; Between data cleaning and inference: Pre-averaging and robust estimators of the efficient price. *Journal of Econometrics*. 194(2):242–262.
- Tao M, Wang Y, Chen X. 2013a; Fast convergence rates in estimating large volatility matrices using high-frequency financial data. *Econometric Theory*. 29(04):838–856.
- Tao M, Wang Y, Zhou HH, et al. 2013b; Optimal sparse volatility matrix estimation for high-dimensional $it\hat{o}$ processes with measurement errors. *The Annals of Statistics*. 41(4):1816–1864.
- Wang Y. 2002; Asymptotic nonequivalence of garch models and diffusions. *The Annals of Statistics*. 30(3):754–783.
- Wang Y, Zou J. 2010; Vast volatility matrix estimation for high-frequency financial data. *The Annals of Statistics*. 38:943–978.
- Xiu D. 2010; Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*. 159(1):235–250.
- Zhang L. 2006; Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*. 12(6):1019–1043.
- Zhang L. 2011; Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*. 160(1):33–47.
- Zhang L, Mykland PA, Ait-Sahalia Y. 2005; A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*. 100(472):1394–1411.

Figure 1.

The relative Frobenius norm error plots of RPRVM, PRVM, PRPRVM, and PPRVM estimators and corresponding thresholding estimators and POET estimators for $p = 512$ and $n = 250, 500, 1000$.

Figure 2.
The spectral norm error plots of the inverse integrated volatility matrix estimators for $p = 512$ and $n = 250, 500, 1000$.

Figure 3.

The spectral norm error plots of the sparse integrated volatility matrix estimators for $p = 512$ and $n = 250, 500, 1000$.

Figure 4.
The scree plots for median eigenvalues of RPRVM and PRVM.

Table 1

The relative Frobenius norm error of RPRVM, PRVM, PRPRVM, and PPRVM estimators and corresponding thresholding estimators and POET estimators for $p = 512$ and $n = 250, 500, 1000$. The standard errors are given in the parentheses.

Projection	Tail	n	Original			Thresholding only			POET	
			RPRVM	PRVM		RPRVM	PRVM		RPRVM	PRVM
Original	Heavy	250	3.40 (0.06)	4.08 (0.07)		3.18 (0.13)	3.50 (0.11)		1.22 (0.06)	1.51 (0.07)
		500	2.90 (0.04)	3.41 (0.04)		2.82 (0.10)	3.10 (0.10)		0.89 (0.04)	1.07 (0.05)
		1000	2.51 (0.03)	2.93 (0.03)		2.47 (0.06)	2.74 (0.07)		0.69 (0.03)	0.82 (0.03)
	Sub-Gaussian	250	3.53 (0.05)	4.09 (0.06)		2.64 (0.08)	2.89 (0.12)		1.38 (0.06)	1.66 (0.07)
		500	3.00 (0.03)	3.41 (0.04)		2.46 (0.04)	2.69 (0.06)		1.03 (0.04)	1.23 (0.04)
		1000	2.59 (0.02)	2.93 (0.03)		2.26 (0.04)	2.47 (0.05)		0.82 (0.03)	0.98 (0.03)
PSD	Heavy	250	3.32 (0.06)	4.07 (0.07)		3.11 (0.13)	3.50 (0.11)		1.20 (0.06)	1.51 (0.07)
		500	2.83 (0.04)	3.40 (0.04)		2.75 (0.09)	3.10 (0.10)		0.87 (0.04)	1.07 (0.05)
		1000	2.44 (0.03)	2.93 (0.03)		2.41 (0.06)	2.74 (0.07)		0.67 (0.03)	0.82 (0.03)
	Sub-Gaussian	250	3.45 (0.05)	4.08 (0.06)		2.63 (0.07)	2.89 (0.11)		1.42 (0.06)	1.67 (0.07)
		500	2.93 (0.03)	3.41 (0.04)		2.44 (0.04)	2.69 (0.06)		1.07 (0.04)	1.24 (0.04)
		1000	2.53 (0.02)	2.92 (0.03)		2.23 (0.03)	2.47 (0.05)		0.86 (0.03)	0.98 (0.03)

Table 2

The MSPEs of sparse integrated volatility matrix estimators from the POET procedures with the RPRVM and PRVM estimators.

MSPE $\times 10^5$					
rank r					
	1	2	3	4	5
RPRVM	1.3960	1.4797	1.5259	1.5575	1.5810
PRVM	2.3339	2.5394	2.6435	2.7068	2.7493