Structured Volatility Matrix Estimation for Non-synchronized High-frequency Financial Data

Jianqing Fan[b,a] and Donggyu Kim[a,c*]

[a]Department of Operations Research and Financial Engineering, Princeton University, USA

[b] School of Economics, Fudan University, Shanghai China

[c]College of Business, Korea Advanced Institute of Science and Technology, Seoul, Korea

April 20, 2019

Abstract

Several large volatility matrix estimation procedures have been recently developed for factor-based Itô processes whose integrated volatility matrix consists of low-rank and sparse matrices. Their performance depends on the accuracy of input volatility matrix estimators. When estimating co-volatilities based on high-frequency data, one of the crucial challenges is non-synchronization for illiquid assets, which makes their co-volatility estimators inaccurate. In this paper, we study how to estimate the large integrated volatility matrix without using co-volatilities of illiquid assets. Specifically, we pretend that the co-volatilities for illiquid assets are missing, and estimate the low-rank matrix using a matrix completion scheme with a structured missing pattern. To further regularize the sparse volatility matrix, we employ the principal orthogonal complement thresholding method (POET). We also investigate the asymptotic properties of the proposed estimation procedure and demonstrate its advantages over using co-volatilities of illiquid assets. The advantages of our methods are also verified by an extensive simulation study and illustrated by high-frequency data for NYSE stocks.

JEL classification: C13, C32, C55.

*corresponding author. Tel: +82 2 958 3448.
E-mail addresses: donggyukim@kaist.ac.kr (D. Kim), jqfan@princeton.edu (J. Fan)

Key words and phrases: Diffusion process, factor model, high-frequency data, low-rank matrix, matrix completion, POET, sparsity.

1 Introduction

High-frequency financial data have provided researchers and practitioners with incredible information to investigate asset pricing and market volatility dynamics. New analytic challenges also arise from analysis of high-frequency financial data. First, due to small market inefficiency such as bid-ask bounce, asymmetric information, latency, and so on, stock prices are contaminated by micro-structural noises. If the micro-structural noises are not accounted for, estimators for integrated volatilities will diverge as the frequency increases (Aït-Sahalia et al., 2005). Second, the observation time points are not synchronized, which makes it hard to estimate co-volatilities, particularly for those illiquid assets. Despite these challenges, several efficient estimation procedures have been developed. Examples 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), pre-averaging realized volatility (PRV) (Christensen et al., 2010; Jacod et al., 2009), kernel realized volatility (KRV) (Barndorff-Nielsen et al., 2008, 2011), quasi-maximum likelihood estimator (QMLE) (Aït-Sahalia et al., 2010; Xiu, 2010), local method of moments (Bibinger et al., 2014), and robust pre-averaging realized volatility (Fan and Kim, 2017).

When estimating co-volatilities, to handle the non-synchronization problem, we often employ some synchronization scheme such as generalized sampling time (Aït-Sahalia et al., 2010), refresh time (Barndorff-Nielsen et al., 2011; Fan et al., 2012), previous tick (Wang and Zou, 2010; Zhang, 2011), and some linear interpolation (Bibinger et al., 2014) schemes. See also Hayashi and Yoshida (2005, 2011); Malliavin and Mancino (2002); Malliavin et al. (2009); Mancino and Sanfelici (2008); Park et al. (2016). These synchronization schemes asymptotically guarantee that the errors coming from the non-synchronized observations can be negligible as the frequency increases. However, for illiquid assets, whose trading frequencies are relatively low, the errors may not be asymptotically negligible, as the refresh

times are too long to be useful so that estimators for co-volatilities can be inaccurate. This generates demand for investigating how to better estimate co-volatilities for illiquid assets. Apparently, we need to appeal to structural aspects of the model.

A commonly used structure to account for cross-sectional dependence is the factor model. It was first used to estimate high-dimensional covariance matrix in Fan et al. (2008) for portfolio allocation and risk management and admits a low-rank plus sparse volatility matrix structure (Fan et al., 2013; Ait-Sahalia and Xiu, 2017; Fan et al., 2016a; Kim et al., 2017; Kong, 2017). When the number of assets is large, the latent factors can be accurately estimated. The performance of these factor-based estimators depends critically on the accuracy of the initial volatility matrix input. However, as discussed above, the co-volatility estimators for illiquid assets are inaccurate, due to relatively long refresh times between any two illiquid assets. On the other hand, the special covariance structure implied by the factor model makes us possible to use the covariance information from liquid blocks to infer about those in illiquid blocks.

How to estimate co-volatilities for illiquid assets, which have serious non-synchronization issue? In this paper, we appeal to the factor structure to infer these co-volatilities. The factor structure implies that the volatility matrix consists of a low-rank covariance matrix induced by the linear combinations of common factors and a sparse covariance matrix induced by idiosyncratic components. We investigate how to estimate the low-rank (or factor) volatility matrix without using estimators for illiquid assets. Due to the low-rankness of the covariance matrix induced by the linear combinations of the common factors, the sub-matrix corresponding to the illiquid assets is spanned by the column space of the remaining low-rank volatility sub-matrices and can be determined analytically from the sub-matrices that involve liquid assets. Thus, the problem of estimating the low-rank volatility matrix is related to the popular matrix completion problem (Candès and Recht, 2009; Koltchinskii et al., 2011), except that the entries (corresponding to the illiquid assets) are not 'missing' at random, but 'missing' (not used due to their inaccuracies) with a structured pattern (Cai et al., 2016). This structured pattern allows us to use the aforementioned analytical formula to estimate the factor-induced volatility submatrix that corresponds to illiquid assets. Then we estimate

the sparse (or idiosyncratic) volatility matrix by subtracting the low-rank volatility estimator from the input volatility matrix estimator and apply the adaptive thresholding scheme to the sparse volatility matrix estimator. The resulting procedure of this kind is called Principal Orthogonal complEment Thresholding (POET) in Fan et al. (2013).

We will investigate the asymptotic behaviors of the proposed estimators for the volatility matrices that correspond to linear combinations of factors, the idiosyncratic components, and the log-returns of assets. We assume that the high-frequency data are contaminated with micro-structural noises. We ideally model the trading volumes of liquid and illiquid assets. We explicitly show when and where the gain can be made by ignoring the co-volatilities of the illiquid assets.

The rest of the paper is organized as follows. Section 2 provides a factor-based diffusion process and data structure and Section 3 reviews the pairwise refresh time scheme and pre-averaging realized volatility estimation method. A large volatility estimation procedure is proposed in Section 4 using matrix completion scheme with the structured missing pattern, whose asymptotic properties are established. The advantages of the proposed method is demonstrated via a simulation study in Section 5 and is illustrated by an application to the NYSE stocks in Section 6. Proofs are collected in Section 7.

2 Model set-up

We first define some notations. For any given vector $\mathbf{a}$, diag($\mathbf{a}$) creates a diagonal matrix using elements of $\mathbf{a}$. For any given $d_1 \times d_2$ matrix $\mathbf{U} = (U_{ij})$,

$$\|\mathbf{U}\|_1 = \max_{1\le j\le d_2} \sum_{i=1}^{d_1} |U_{ij}|, \quad \|\mathbf{U}\|_\infty = \max_{1\le i\le d_1} \sum_{j=1}^{d_2} |U_{ij}|, \quad \text{and} \quad \|\mathbf{U}\|_{\max} = \max_{i,j} |U_{ij}|.$$

Matrix spectral norm $\|\mathbf{U}\|_2$ is the largest eigenvalue of $\mathbf{U}\mathbf{U}^\top$, the Frobenius norm of $\mathbf{U}$ is $\|\mathbf{U}\|_F = \sqrt{\text{tr}(\mathbf{U}^\top\mathbf{U})}$. $\mathbf{U}_{IJ}$ denotes the sub-matrix of $\mathbf{U}$ formed by rows and columns whose indices are in I and J, respectively, where I and J are subsets of $\{1,\ldots,d_1\}$ and $\{1,\ldots,d_2\}$, respectively. We will use C to denote a generic constant whose value is free of n and p and

may change from occurrence to occurrence.

Let $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^\top$ be the vector of true log-prices at time t. We assume that the log-prices of assets follow a continuous-time diffusion model. In economic and financial studies, the approximate factor model is widely employed to account for the effect of macro-economic factors and market factors such as sector and industry classification, firm size, price to book ratios, etc (Bai and Ng, 2002; Chamberlain and Rothschild, 1982; Fama and French, 1992; Fan et al., 2016a; Ait-Sahalia and Xiu, 2017). In light of these, we employ the factor-based diffusion model

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

where $\boldsymbol{\mu}(t) = (\mu_1(t), \ldots, \mu_p(t))^\top$ is a drift vector, $\boldsymbol{\vartheta}(t)$ is a $r \times p$ matrix, $\boldsymbol{\sigma}(t)$ is a $p \times p$ matrix, $\mathbf{W}_t^*$ and $\mathbf{W}_t$ are independent r-dimensional and p-dimensional Brownian motions, respectively. 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. Note that r is the number of latent factors. The instantaneous (or spot) volatility matrix of the log-prices $\mathbf{X}(t)$ in (2.1) is

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

The parameter of interest is the integrated volatility matrix over time $[0,1]$

$$\begin{aligned} \boldsymbol{\Gamma} &= \int_0^1 \boldsymbol{\gamma}(t)dt \\ &= \int_0^1 \boldsymbol{\vartheta}^\top(t)\boldsymbol{\vartheta}(t)dt + \int_0^1 \boldsymbol{\sigma}^\top(t)\boldsymbol{\sigma}(t)dt \\ &= \boldsymbol{\Theta} + \boldsymbol{\Sigma}. \end{aligned} \tag{2.2}$$

The matrix $\boldsymbol{\Theta}$ in (2.2) accounts for the factor influence on the volatility matrix. In this paper, we assume that the rank, r, of $\boldsymbol{\Theta}$ is fixed and finite. Additionally, we impose some sparse structure on the idiosyncratic volatility matrix $\boldsymbol{\Sigma}$ (see Section 4). Thus, the integrated volatility matrix $\boldsymbol{\Gamma}$ has the low-rank plus sparse structure which is widely used in analyzing

large covariance or volatility matrices (Fan et al., 2013, 2016b; Ait-Sahalia and Xiu, 2017; Kim et al., 2017; Kong, 2017).

Unfortunately, in the high-frequency finance, we cannot observe the true log-prices due to the micro-structural noises caused by small market inefficiencies, for example, asymmetric information, bid-ask bounce, and latency. We also encounter the so-called non-synchronization problem that transactions for different assets occur at distinct times, and the observation time points are not synchronized. To model these stylized features, in the high-frequency finance, it is usually assumed that the observed price $Y_i(t_{i,k})$ has an additive noise as follows:

$$Y_i(t_{i,k}) = X_i(t_{i,k}) + \epsilon_i(t_{i,k}) \quad \text{for } i = 1, \ldots, p, k = 0, \ldots, n_i, \tag{2.3}$$

where $\epsilon_i(t_{i,k}), i = 1, \ldots, p, k = 0, \ldots, n_i$, are independent noises with mean zero and variance η_{ii} and p is the number of assets. Furthermore, we observe that the numbers, $n_1, \ldots, n_p$, of high-frequency observations are heterogeneous. For the simplicity, we assume that there are two sub-groups of stocks which have high trading volumes (liquid assets) and low trading volumes (illiquid assets) as follows:

$$\mathcal{H} = \{i \in \{1, \ldots, p\}, n_i \asymp n\} \quad \text{and} \quad \mathcal{L} = \{i \in \{1, \ldots, p\}, n_i \asymp n^a\}, \tag{2.4}$$

where $a < 1$ and $\mathcal{H} \cup \mathcal{L} = \{1, \ldots, p\}$. Their cardinalities are $|\mathcal{H}| = p_1$ and $|\mathcal{L}| = p_2$. Then, without loss of generality, we can rearrange the integrated volatility matrix $\mathbf{\Gamma}$ as follows:

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} \end{pmatrix},$$

where $\mathbf{\Gamma}_{11} = \mathbf{\Gamma}_{\mathcal{H}\mathcal{H}}$, $\mathbf{\Gamma}_{12} = \mathbf{\Gamma}_{\mathcal{H}\mathcal{L}}$, $\mathbf{\Gamma}_{21} = \mathbf{\Gamma}_{\mathcal{L}\mathcal{H}}$, and $\mathbf{\Gamma}_{22} = \mathbf{\Gamma}_{\mathcal{L}\mathcal{L}}$. Note that the sub-matrices have the low-rank plus sparse structure as follows:

$$\mathbf{\Gamma}_{ij} = \mathbf{\Theta}_{ij} + \mathbf{\Sigma}_{ij} \quad \text{for } i = 1, 2, j = 1, 2,$$

when we use the following partitions:

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\Theta}_{11} & \boldsymbol{\Theta}_{12} \\ \boldsymbol{\Theta}_{21} & \boldsymbol{\Theta}_{22} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Due to the errors coming from non-synchronized observation time points, co-volatility estimators are less accurate especially for the low trading volume set $\mathcal{L}$. That is, estimators for co-volatilities of $\boldsymbol{\Gamma}_{22}$ are less accurate than those of other blocks $\boldsymbol{\Gamma}_{11}$ and $\boldsymbol{\Gamma}_{12}$. In light of this, in this paper, we study how to estimate the integrated volatility matrix $\boldsymbol{\Gamma}$ without estimating the off-diagonal elements of $\boldsymbol{\Gamma}_{22}$.

3 Co-volatility estimation

3.1 Pairwise refresh method

To handle the non-synchronization problem, we can use synchronization schemes such as generalized sampling time (Aït-Sahalia et al., 2010), refresh time (Barndorff-Nielsen et al., 2011; Fan et al., 2012), and previous tick (Wang and Zou, 2010; Zhang, 2011) schemes, or some linear interpolation scheme (Bibinger et al., 2014). There are estimation procedures which do not require to align data (Hayashi and Yoshida, 2005, 2011; Malliavin and Mancino, 2002; Malliavin et al., 2009; Mancino and Sanfelici, 2008; Park et al., 2016). One way to utilize the data efficiently is to apply the pairwise refresh time scheme to estimate co-volatility. Given the k-th refresh time, the $(k+1)$-th refresh time is the minimum calendar time needed for both stock to be traded at least once. The formal definition is as follows.

Definition 1. *Let $\{t_{ij}\}_{j=1}^{n_i}$ be the calendar times where the i^{th} stock is traded as in* (2.3)*. The first refresh time for the i-th and j-th assets is defined as $\tau_{ij,1} = \max\{t_{i,1}, t_{j,1}\}$. The subsequent refresh times are*

$$\tau_{ij,k+1} = \max\{t_{i,N_i(\tau_{ij,k})+1}, t_{j,N_j(\tau_{ij,k})+1}\},$$

where $N_i(t)$ is the number of observations in the i-th asset made up to time t.

With the refresh time scheme, for the i-th asset, we select any observation, $Y_i(t_{i,k})$, for $t_{i,k}$ between $\tau_{ij,k-1}$ and $\tau_{ij,k}$, to be paired or synchronized with $Y_j(t_{j,k})$ with $t_{j,k}$ chosen similarly, for computing the co-volatility of asset i and asset j. Let $\bar{n}_{ij}$ be the number of such synchronized observations for the i-th and j-th assets. Then $\bar{n}_{ij} \leq \min(n_i, n_j)$ and $\bar{n}_{ii} = n_i$.

3.2 Pre-averaging realized volatility estimation

To handle the micro-structural noise, several estimation methods have been developed and the error from the noise can be removed effectively (see Aït-Sahalia et al. (2010); Barndorff-Nielsen et al. (2008, 2011); Bibinger et al. (2014); Christensen et al. (2010); Fan and Wang (2007); Jacod et al. (2009); Xiu (2010); Zhang et al. (2005); Zhang (2006, 2011)). In this paper, we use the pre-averaging realized volatility estimation scheme (Christensen et al., 2010; Jacod et al., 2009).

Definition 2 (Christensen et al. (2010); Jacod et al. (2009))**.** *For the pairwise refresh time, $\{\tau_{ij,k}\}_{k=1}^{n}$ with $n = \bar{n}_{ij}$, the pre-averaging realized volatility (PRV) estimator is given by*

$$\widehat{\Gamma}_{ij} = \frac{1}{\psi K} \sum_{k=1}^{n-K+1} \left\{ Z_i(\tau_{ij,k}) Z_j(\tau_{ij,k}) - \varsigma \, \widehat{\eta}_{ij} \, \mathbf{1}(i=j) \right\},$$

where $\psi = \int_0^1 g^2(t)dt$,

$$\begin{aligned}
\widehat{\eta}_{ii} &= \frac{1}{2n_i} \sum_{k=1}^{n_i} \left\{ Y_i(t_{i,k}) - Y_i(t_{i,k-1}) \right\}^2, \\
Z_i(\tau_{ij,k}) &= \sum_{l=1}^{K-1} g\left(\frac{l}{K}\right) \left\{ Y_i(\tau_{ij,k+l}) - Y_i(\tau_{ij,k+l-1}) \right\}, \\
\varsigma &= \sum_{l=0}^{K-1} \left\{ g\left(\frac{l}{K}\right) - g\left(\frac{l+1}{K}\right) \right\}^2 = O\left(\frac{1}{K}\right),
\end{aligned}$$

$K = Cn^{1/2}$ is a bandwidth parameter for some constant C free of n and p, and $g(\cdot)$ is a weight function satisfying that g is continuous and piecewise continuously differentiable with

a piecewise Lipschitz derivative g' and satisfies $g(0) = g(1) = 0$.

Remark 1. The bias correction term $\widehat{\eta}_{ij}$ is required to obtain the optimal convergence rate $n^{-1/4}$ with the presence of the micro-structural noise. In this paper, we simply assume that the micro-structural noises are independent and so their diagonal parts are only required to be estimated. When they have some correlation structure, we may need to estimate the off-diagonal parts η_{ij} for $i \neq j$. When it comes to constructing estimation procedures for co-volatility part η_{ij}, due to the non-synchronization problem, we need to define the correlation structure carefully, and the estimation procedures are depending on the correlation structure. Kim et al. (2016) discussed and studied this issue. Fortunately, as long as we can estimate the co-volatilities well, theoretical results obtained in this paper will be the same. Thus, to focus on solving the non-synchronization problem, we simply assume that the micro-structural noises are independent.

To investigate the large volatility matrices, we need the sub-Gaussian concentration inequality

$$\Pr\left(|\widehat{\Gamma}_{ij} - \Gamma_{ij}| \geq C_m \sqrt{\log p / \bar{n}_{ij}^{1/2}}\right) \leq p^{-m},$$

where C_m is some constant depending only on given constant m. With mild conditions, Kim and Wang (2016) studied its sub-Gaussian concentration inequality. We will utilize their result.

Assumption 1.

(1) There are some fixed constants C_μ and C_σ such that, almost surely,

$$\max_{1\leq i\leq p} \max_{0\leq t\leq 1} |\mu_i(t)| \leq C_\mu, \qquad \max_{1\leq i\leq p} \max_{0\leq t\leq 1} \gamma_{ii}(t) \leq C_\sigma;$$

(2) $\epsilon_i(t_{i,k})$ and $\mathbf{X}(t)$ are independent. For each i, $\epsilon_i(t_{i,k})$, $k = 0, \ldots, n_i$, have sub-Gaussian distributions;

(3) The observation time points are independent with log-stock price processes $\mathbf{X}(t)$ and micro-structural noises $\epsilon_i(t_{i,k})$'s, and the pairwise refresh time points $\tau_{ij,k}$ satisfy

$\max_{1\le i,j\le p}\max_{1\le k\le \bar{n}_{ij}}(\tau_{ij,k}-\tau_{ij,k-1})\bar{n}_{ij}\le C_\tau$ *a.s. for some generic constant* C_τ *free of* n *and* p.

Remark 2. Assumption 1 is usually assumed to obtain the sub-Gaussian concentration inequality which plays an important role in the high-dimensional inferences (Tao et al., 2013; Kim and Wang, 2016). Recently, Fan and Kim (2017) proposed the robust pre-averaging realized volatility which can obtain the sub-Gaussian concentration inequality with only the finite fourth moment condition. The sub-Gaussian conditions Assumption 1 (1)–(2) can be relaxed by employing the robust pre-averaging realized volatility. Assumption 1 (3) indicates that the time intervals for each pair have the order $\bar{n}_{ij}^{-1}$ which goes to zero as the sample size goes to infinity.

Propostion 3.1 (Theorem 1 (Kim and Wang, 2016))**.** *Under the models* (2.1) *and* (2.3)*, if Assumption 1 is met, then the pre-averaging realized volatility estimator* $\widehat{\Gamma}_{ij}$ *in Definition 2 has the following sub-Gaussian concentration:*

$$\Pr\left(|\widehat{\Gamma}_{ij}-\Gamma_{ij}|\ge x\right)\le \vartheta_1\exp\left(-\sqrt{\bar{n}_{ij}}x^2/\vartheta_2\right), \tag{3.1}$$

where x *is a positive number in a neighbor of* 0*, and* ϑ_1 *and* ϑ_2 *are generic constants free of* n *and* p.

We need only the input volatility estimator that satisfies the sub-Gaussian concentration inequality (3.1) in order to investigate the asymptotic behavior of the proposed estimation procedure. Thus, we can use any other estimation procedure satisfying (3.1). For example, multi-scale realized volatility (Zhang, 2006, 2011) and robust pre-averaging realized volatility (Fan and Kim, 2017) can be used. In the numerical analysis, we use the pre-averaging realized volatility matrix (PRVM) estimation procedure in Definition 2 with $K=n^{1/2}$ and $g(x)=x\wedge(1-x)$.

4 Large volatility matrix estimation

4.1 Low-rank volatility matrix estimation

Several large volatility matrix estimation procedures have been developed based on the factor model (Fan et al., 2016a; Ait-Sahalia and Xiu, 2017; Kim et al., 2017; Kong, 2017). Their performances may depend on the accuracy of the input volatility matrix estimator $\widehat{\mathbf{\Gamma}}$. As discussed in Section 2, when it comes to estimating co-volatilities in high-frequency finance, one of the crucial issues is the non-synchronization problem. We use the pairwise refresh time defined in Definition 1 in order to utilize the information efficiently. Then when estimating co-volatilities for liquid assets $\mathcal{H}$, the estimation errors coming from the non-synchronized observations can be small. Thus, we can estimate the co-volatilities well in the corresponding block $\mathbf{\Gamma}_{11}$. On the other hand, when estimating co-volatilities for illiquid assets $\mathcal{L}$, it is hard to expect that the estimated co-volatilities are accurate due to the errors coming from non-synchronized observation time points. The intervals for refresh time can be so large that the approximation errors are too big for applications. In this section, we investigate how to estimate the low-rank (or factor) volatility matrix $\mathbf{\Theta}$ without estimating the co-volatilities for illiquid assets $\mathcal{L}$.

In order to investigate the effect of the non-synchronization problem in estimating co-volatilities, we assume that the number of the synchronized time points is

$$\bar{n}_{ij} = c \min\left\{ \left(\frac{n_i + n_j}{2} \right)^b, n_i, n_j \right\} \quad \text{for } i \neq j, \tag{4.1}$$

where some generic constant $c \leq 1$ and $b \in (a, 1)$ with a defined in (2.4). In literature, researchers usually assume that $b = 1$ and $\bar{n}_{ij} = c \min(n_i, n_j)$. However, this is too optimistic due to lost of data in the synchronization process and hence we will assume $b < 1$. Combining

(2.4) and (4.1), we have

$$\begin{cases} \bar{n}_{ij} \asymp n^b & \text{if } i,j \in \mathcal{H} \\ \bar{n}_{ij} \asymp n^a & \text{if } i \in \mathcal{H}, j \in \mathcal{L} \\ \bar{n}_{ij} \asymp n^{ab} & \text{if } i,j \in \mathcal{L}, \end{cases}$$

where $\bar{n}_{ij} = \bar{n}_{ji}$. The above formula is a reasonable model, since for the synchronization between liquid and illiquid assets, it is reasonable to assume that we are able to observe the liquid assets around each observation time point of illiquid assets so that $\bar{n}_{ij} \asymp n^a$. On the other hand, for the synchronization of similar liquidity assets (liquid-liquid or illiquid-illiquid), there is some cost to align the data, which is mathematically expressed by $b \in (a,1)$. Thus, under the assumption (4.1), the estimators for the off-diagonal elements of $\mathbf{\Gamma}_{22}$ have slower convergence rates.

To account for the common factors in the financial market, we assume that the integrated volatility matrix $\mathbf{\Gamma}$ consists of the low-rank and sparse matrices with the block structure as follows:

$$\mathbf{\Gamma} \;=\; \mathbf{\Theta} + \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Theta}_{11} & \mathbf{\Theta}_{12} \\ \mathbf{\Theta}_{21} & \mathbf{\Theta}_{22} \end{pmatrix} + \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}.$$

The volatility matrix $\mathbf{\Sigma}$ of the idiosyncratic component is sparse in the sense that it satisfies

$$\max_{1 \le j \le p} \sum_{i=1}^{p} |\Sigma_{ij}|^q (\Sigma_{ii}\Sigma_{jj})^{(1-q)/2} \le M_\sigma s(p) \text{ a.s.}, \tag{4.2}$$

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. Here we define $0^0 = 0$. For the exact sparse matrix, that is, $q = 0$, when Σ_{ii} is bounded from below, the sparsity level $s(p)$ measures the maximum number of non-vanishing elements in each row of the idiosyncratic volatility matrix $\mathbf{\Sigma}$.

As discussed before, due to the non-synchronization problem, the estimators for the off-diagonal elements of $\mathbf{\Gamma}_{22}$ may not be accurate. With the inaccurate estimator, when we

apply the POET procedure (Fan et al., 2013) to estimating the low-rank volatility matrix $\mathbf{\Theta}$, the resulting estimator may have a poor asymptotic behavior due to the inaccuracy of the input volatility matrix. The simulation study supports this (see Section 5). To avoid this problem, we do not use the illiquid asset information for estimating $\mathbf{\Theta}_{22}$, but get a better estimator for $\mathbf{\Theta}_{22}$ using the low-rank structure of $\mathbf{\Theta}$.

Note that $\mathbf{\Theta}_{11}$ is a $p_1 \times p_1$ integrated volatility matrix of p_1 liquid assets. Let $\lambda_i, i = 1, \ldots, r$, be the eigenvalues of $\mathbf{\Theta}_{11}$ with decreasing order and $\mathbf{Q} \in \mathbb{R}^{p_1 \times r}$ be the matrix of their associated eigenvectors. When the rank of $\mathbf{\Theta}_{11}$ is r, which is the number of the latent factors, it admits the spectral decomposition:

$$\mathbf{\Theta}_{11} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top, \quad \text{where} \quad \mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_r).$$

Since $\mathbf{\Theta}$ has a rank r, $\mathbf{\Theta}_{22}$ must be the linear combinations of the columns spanned by $\mathbf{\Theta}_{21}$. It can easily be shown that

$$\mathbf{\Theta}_{22} = \mathbf{\Theta}_{21}\mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^\top\mathbf{\Theta}_{12}; \tag{4.3}$$

see Proposition 1 in Cai et al. (2016). Also, columns of $\mathbf{\Theta}_{12}$ are linear combinations of $\mathbf{\Theta}_{11}$ as follows:

$$\mathbf{\Theta}_{12} = \mathbf{\Theta}_{11}\mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^\top\mathbf{\Theta}_{12} = \mathbf{Q}\mathbf{Q}^\top\mathbf{\Theta}_{12}. \tag{4.4}$$

Thus, as long as we have well-performing estimators for $\mathbf{\Theta}_{11}$ and $\mathbf{\Theta}_{12}$, we can construct the low-rank volatility matrix $\mathbf{\Theta}$ using the relationship in (4.3) and (4.4). Identity (4.4) will be used below to ensure that the rank of empirically constructed $\widehat{\mathbf{\Theta}}$ has the rank r. See Remark 3.

For any estimators $\widehat{\Gamma}_{ij}$ for Γ_{ij}, let

$$\widehat{\mathbf{\Gamma}}_{11} = (\widehat{\Gamma}_{ij})_{i,j\in\mathcal{H}}, \quad \widehat{\mathbf{\Gamma}}_{22} = (\widehat{\Gamma}_{ij})_{i,j\in\mathcal{L}}, \quad \widehat{\mathbf{\Gamma}}_{12} = (\widehat{\Gamma}_{ij})_{i\in\mathcal{H},j\in\mathcal{L}}, \quad \text{and} \quad \widehat{\mathbf{\Gamma}}_{21} = \widehat{\mathbf{\Gamma}}_{12}^\top.$$

The corresponding true volatility sub-matrices $\mathbf{\Gamma}_{11}$ and $\mathbf{\Gamma}_{12}$ have the low-rank plus sparse structure. To estimate the latent low-rank volatility sub-matrices, $\mathbf{\Theta}_{11}$, $\mathbf{\Theta}_{12}$, and $\mathbf{\Theta}_{22}$, we employ the POET procedure, and then use the relationship (4.3) and (4.4) to construct the

low-rank volatility matrix $\mathbf{\Theta}$. For example, let the singular value decompositions of $\widehat{\mathbf{\Gamma}}_{11}$ and $\widehat{\mathbf{\Gamma}}_{12}$ be

$$\widehat{\mathbf{\Gamma}}_{11} = \sum_{k=1}^{p_1} \widehat{\lambda}_k \widehat{\mathbf{q}}_k \widehat{\mathbf{q}}_k^\top \quad \text{and} \quad \widehat{\mathbf{\Gamma}}_{12} = \sum_{k=1}^{p_1 \wedge p_2} \widehat{\xi}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{u}}_k^\top,$$

where $\widehat{\lambda}_k$ and $\widehat{\xi}_k$ are the k-th largest singular values of $\widehat{\mathbf{\Gamma}}_{11}$ and $\widehat{\mathbf{\Gamma}}_{12}$, respectively, $\widehat{\mathbf{q}}_k$ are the singular vectors (eigenvectors) corresponding to $\widehat{\lambda}_k$, and $\widehat{\mathbf{u}}_k$ and $\widehat{\mathbf{v}}_k$ are the left and right singular vectors corresponding to $\widehat{\xi}_k$. Using the plug-in procedure, we estimate the low-rank volatility sub-matrices $\mathbf{\Theta}_{11}$ and $\mathbf{\Theta}_{12}$ by

$$\widehat{\mathbf{\Theta}}_{11} = \widehat{\mathbf{Q}} \widehat{\mathbf{\Lambda}} \widehat{\mathbf{Q}}^\top \qquad \text{and} \qquad \widetilde{\mathbf{\Theta}}_{12} = \sum_{k=1}^{r} \widehat{\xi}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{u}}_k^\top,$$

respectively, where $\widehat{\mathbf{Q}} = (\widehat{\mathbf{q}}_1, \ldots, \widehat{\mathbf{q}}_r)$ and $\widehat{\mathbf{\Lambda}} = \mathrm{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r)$. Under the pervasive and incoherence conditions (Assumption 2(d)–(e)), they will be shown to have good asymptotic performances. The liquid asset block estimator $\widehat{\mathbf{\Theta}}_{11}$ is the most accurate estimator and will be used as the pivotal estimator. We estimate the other blocks, $\mathbf{\Theta}_{12}$ and $\mathbf{\Theta}_{22}$, using the relationship (4.3) and (4.4) as follows:

$$\widehat{\mathbf{\Theta}}_{22} = \widehat{\mathbf{\Theta}}_{21} \widehat{\mathbf{Q}} \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{Q}}^\top \widehat{\mathbf{\Theta}}_{12} \qquad \text{and} \qquad \widehat{\mathbf{\Theta}}_{12} = \widehat{\mathbf{\Theta}}_{11} \widehat{\mathbf{Q}} \widehat{\mathbf{\Lambda}}^{-1} \widehat{\mathbf{Q}}^\top \widetilde{\mathbf{\Theta}}_{12} = \widehat{\mathbf{Q}} \widehat{\mathbf{Q}}^\top \widetilde{\mathbf{\Theta}}_{12}.$$

See Remark 3 for the reason why we do not use directly $\widetilde{\mathbf{\Theta}}_{12}$ in the both expression above. Combining the low-rank volatility sub-matrix estimators, we estimate the low-rank volatility matrix estimator by

$$\widehat{\mathbf{\Theta}} = \begin{pmatrix} \widehat{\mathbf{\Theta}}_{11} & \widehat{\mathbf{\Theta}}_{12} \\ \widehat{\mathbf{\Theta}}_{21} & \widehat{\mathbf{\Theta}}_{22} \end{pmatrix}, \tag{4.5}$$

where $\widehat{\mathbf{\Theta}}_{21} = \widehat{\mathbf{\Theta}}_{12}^\top$. We call it the structured low-rank volatility matrix estimator.

Remark 3. The other possible estimator of $\mathbf{\Theta}$ is

$$\widehat{\mathbf{\Theta}}^{alt} = \begin{pmatrix} \widehat{\mathbf{\Theta}}_{11} & \widetilde{\mathbf{\Theta}}_{12} \\ \widetilde{\mathbf{\Theta}}_{21} & \widehat{\mathbf{\Theta}}_{22} \end{pmatrix}, \tag{4.6}$$

which has the same element-wise convergence rate of the proposed estimator in (4.5). However, for the finite sample, we cannot guarantee that the rank of $\widehat{\mathbf{\Theta}}^{alt}$ is r. This is because the colomns of $\widetilde{\mathbf{\Theta}}_{12}$ is not necessary in the space spanned by the columns of $\widehat{\mathbf{\Theta}}_{11}$. In contrast, by the construction of $\widehat{\mathbf{\Theta}}_{12}$, the structured low-rank volatility estimator $\widehat{\mathbf{\Theta}}$ has the rank r, which is one of the desired properties. For the same reason, we used $\widehat{\mathbf{\Theta}}_{12}$ instead of $\widetilde{\mathbf{\Theta}}_{12}$ in constructing $\widehat{\mathbf{\Theta}}_{22}$. The simulation study in Section 5 indicates that $\widehat{\mathbf{\Theta}}$ outperforms $\widehat{\mathbf{\Theta}}^{alt}$.

To investigate the asymptotic behavior of the low-rank volatility matrix estimator $\widehat{\mathbf{\Theta}}$, we make several technical conditions.

Assumption 2.

(a) The ranks of $\mathbf{\Theta}_{11}$ and $\mathbf{\Theta}$ are the same;

(b) There are some deterministic sequences $\beta_{1,n}$ and $\beta_{2,n}$ such that, with probability greater than $1-\delta$,

$$\|\widehat{\mathbf{\Gamma}}_{11} - \mathbf{\Gamma}_{11}\|_{\max} \leq \beta_{1,n} = o(1), \qquad \|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Gamma}_{12}\|_{\max} \leq \beta_{2,n} = o(1),$$

and $\beta_{1,n} \leq \beta_{2,n}$;

(c) The sparsity level diverges slowly such that $s(p)/\sqrt{p_1 \wedge p_2} = o(1)$;

(d) Let $D_\lambda = \min\{\lambda_i - \lambda_{i+1} : 1 \leq i \leq r\}$ and $D_\xi = \min\{\xi_i - \xi_{i+1} : 1 \leq i \leq r\}$, and there are some fixed constants $c_1, \ldots, c_4$ such that $\lambda_1/D_\lambda + p_1 M_\sigma/D_\lambda + \xi_1/D_\xi + \sqrt{p_1 p_2} M_\sigma/D_\xi \leq c_1$, $\xi_1/D_\lambda \leq c_2\sqrt{p_2/p_1}$, $D_\xi \geq c_3\sqrt{p_1 p_2}$, and $D_\lambda \geq c_4 p_1$ almost surely, where ξ_i's are singular values of $\mathbf{\Theta}_{12}$ with decreasing order;

(e) For some fixed constants c_5, c_6, and c_7, we have almost surely

$$\frac{p_1}{r}\max_{1\leq i\leq p_1}\sum_{j=1}^{r} q_{ij}^2 \leq c_5, \qquad \frac{p_1}{r}\max_{1\leq i\leq p_1}\sum_{j=1}^{r} v_{ij}^2 \leq c_6, \qquad \frac{p_2}{r}\max_{1\leq i\leq p_2}\sum_{j=1}^{r} u_{ij}^2 \leq c_7,$$

where $\mathbf{Q} = (q_{ij})_{1\leq i\leq p_1, 1\leq j\leq r}$ is the eigenvector matrix of $\mathbf{\Theta}_{11}$, and $\mathbf{V} = (v_{ij})_{1\leq i\leq p_1, 1\leq j\leq r}$ and $\mathbf{U} = (u_{ij})_{1\leq i\leq p_2, 1\leq j\leq r}$ are the left and right singular vector matrices of $\mathbf{\Theta}_{12}$.

Remark 4. Assumption 2(a) indicates that the liquid-liquid block $\mathbf{\Theta}_{11}$ has the full information of the low-rank volatility matrix $\mathbf{\Theta}$ and the liquid-illiquid block $\mathbf{\Theta}_{12}$ provides the linear relationship between $\mathbf{\Theta}_{11}$ and $\mathbf{\Theta}_{22}$. This assumption allows us to use the accurate estimator $\widehat{\mathbf{\Theta}}_{11}$ as the pivotal estimator. The common factor affects on the whole stock prices and so the corresponding volatility matrix $\mathbf{\Theta}$ is dense. This implies that eigenvalues of $\mathbf{\Theta}$ increases with the p order. Thus, the so-called pervasive condition (Assumption 2(d)) is reasonable to impose on the factor volatility matrix $\mathbf{\Theta}$. Assumption 2(e) is called the incoherence condition which is widely used in analyzing low-rank matrices (see Candès and Recht (2009); Fan et al. (2016b)). This technical condition allows us to analyze the element-wise asymptotic behavior of the factor volatility matrix $\mathbf{\Theta}$.

The following theorem shows the element-wise convergence rate of the structured low-rank volatility matrix estimator $\widehat{\mathbf{\Theta}}$.

Theorem 4.1. *Under the models* (2.1) *and* (2.3)*, if Assumption 2 and the sparse condition* (4.2) *are met, then the structured low-rank volatility matrix estimator in* (4.5) *has for large* n*, with probability greater than* $1-\delta$*,*

$$\|\widehat{\mathbf{\Theta}}_{11}-\mathbf{\Theta}_{11}\|_{\max} \le C\left\{\beta_{1,n}+M_\sigma\frac{s(p)}{p_1}\right\}, \tag{4.7}$$

$$\|\widehat{\mathbf{\Theta}}_{12}-\mathbf{\Theta}_{12}\|_{\max} \le C\left\{\beta_{2,n}+M_\sigma\frac{s(p)}{p_1\wedge p_2}\right\}, \tag{4.8}$$

$$\|\widehat{\mathbf{\Theta}}_{22}-\mathbf{\Theta}_{22}\|_{\max} \le C\left\{\beta_{2,n}+M_\sigma\frac{s(p)}{p_1\wedge p_2}\right\}. \tag{4.9}$$

Remark 5. Under the assumption (4.1), Proposition 3.1 shows that the pre-averaging realized volatility estimator have, with probability greater than $1-p^{-1}$, $\beta_{1,n}=C\sqrt{\log p/n^{b/2}}$ and $\beta_{2,n}=C\sqrt{\log p/n^{a/2}}$. In the financial market, the numbers of stocks in the high trading volume and low trading volume, $\mathcal{H}$ and $\mathcal{L}$, are comparable, and so $p_1 \asymp p_2$. Then Theorem 4.1 shows that the low-rank volatility matrix estimator $\widehat{\mathbf{\Theta}}$ has, with probability greater than $1-p^{-1}$,

$$\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_{\max} \le C\left\{\sqrt{\log p/n^{a/2}}+M_\sigma\frac{s(p)}{p}\right\}.$$

On the other hand, when estimating the low-rank volatility matrix $\mathbf{\Theta}$ using the POET

procedure (Fan and Kim, 2017; Fan et al., 2013), we have, with probability greater than $1-p^{-1}$,

$$\|\widehat{\mathbf{\Theta}}_{POET}-\mathbf{\Theta}\|_{\max}\leq C\left\{\sqrt{\log p/n^{ab/2}}+M_\sigma\frac{s(p)}{p}\right\},$$

where $\widehat{\mathbf{\Theta}}_{POET}$ is the low-rank volatility matrix estimator calculated from the POET procedure. Due to the inaccurate estimator for the off-diagonal elements of $\mathbf{\Gamma}_{22}$, $\widehat{\mathbf{\Theta}}_{POET}$ has the term $\sqrt{\log p/n^{ab/2}}$.

4.2 Sparse volatility matrix estimation

We can estimate the sparse (or idiosyncratic) volatility matrix using some thresholding procedures. For the general sparse structure (4.2), we still need to estimate the off-diagonal elements of $\mathbf{\Gamma}_{22}$, which causes slower convergence rates for the sparse volatility matrix $\mathbf{\Sigma}$. To avoid this, we need to impose more structure on $\mathbf{\Gamma}_{22}$ that it is a low-rank plus a diagonal matrix (Fan et al., 2008). That is, the idiosyncratic risks for the illiquid assets are uncorrelated and satisfies

$$\mathbf{\Sigma}_{22}=\operatorname{diag}\left((\Sigma_{ii})_{i\in\mathcal{L}}\right). \tag{4.10}$$

Based on the sparse conditions (4.2) and (4.10), we estimate the sparse idiosyncratic volatility matrix $\mathbf{\Sigma}$ by letting $\widetilde{\Sigma}_{ij}=\widehat{\Gamma}_{ij}-\widehat{\Theta}_{ij}$ and

$$\widehat{\Sigma}_{ij}=\begin{cases}\widetilde{\Sigma}_{ij}\vee 0, & \text{if } i=j\\ s_{ij}(\widetilde{\Sigma}_{ij})\mathbf{1}(|\widetilde{\Sigma}_{ij}|\geq\varpi_{ij}), & \text{if } i\neq j \text{ and } (i,j)\notin\{(l,k):l,k\in\mathcal{L}\}\\ 0, & \text{if } i\neq j \text{ and } (i,j)\in\{(l,k):l,k\in\mathcal{L}\},\end{cases}$$

where the adaptive thresholding level $\varpi_{ij}=\{(\varpi_{1,n}-\varpi_{2,n})\mathbf{1}(i,j\in\mathcal{H})+\varpi_{2,n}\}\sqrt{(\widetilde{\Sigma}_{ii}\vee 0)(\widetilde{\Sigma}_{jj}\vee 0)}$, and $s_{ij}(\cdot)$ satisfies that $|s_{ij}(x)-x|\leq\varpi_{ij}$. The shrinkage function $s_{ij}(x)$ includes the useful examples such as the soft thresholding function $s_{ij}(x)=x-sign(x)\varpi_{ij}$ and the hard thresholding function $s_{ij}(x)=x$. The tuning parameters $\varpi_{1,n}$ and $\varpi_{2,n}$ will be specified in Theorem 4.2.

With the structured low-rank volatility matrix estimator $\widehat{\mathbf{\Theta}} = (\widehat{\Theta}_{ij})_{1\le i,j\le p}$ in (4.5) and the sparse volatility matrix estimator $\widehat{\mathbf{\Sigma}} = (\widehat{\Sigma}_{ij})_{1\le i,j\le p}$, we estimate the integrated volatility matrix $\mathbf{\Gamma}$ by

$$\widetilde{\mathbf{\Gamma}} = \widehat{\mathbf{\Theta}} + \widehat{\mathbf{\Sigma}}.$$

We call it the structured POET (SPOET) estimator.

To investigate the asymptotic behavior of the SPOET, we make the following technical conditions.

Assumption 3.

(a) We have, with probability greater than $1-\delta$,

$$\max_{i\in\mathcal{L}} |\widehat{\Gamma}_{ii} - \Gamma_{ii}| \le \beta_{2,n} = o(1) \quad \textit{and} \quad M_\sigma \le C;$$

(b) (Pervasive condition) There are some fixed constants c_8 *and* c_9 *such that* $\lambda_r(\mathbf{\Theta}) \ge c_8 p$ *and* $\lambda_1(\mathbf{\Theta})/\lambda_r(\mathbf{\Theta}) \le c_8$ *almost surely, where* $\lambda_k(\mathbf{\Theta})$ *is the* k*-th largest eigenvalue of* $\mathbf{\Theta}$.

The following theorem shows the convergence rate of the proposed SPOET estimator.

Theorem 4.2. *Under the models* (2.1) *and* (2.3)*, assume that Assumptions 2–3, the sparse conditions* (4.2) *and* (4.10)*, and* (4.7)-(4.9) *are met. Take* $\varpi_{1,n} = C_{1,\varpi}(\beta_{1,n} + M_\sigma s(p)/p_1)$ *and* $\varpi_{2,n} = C_{2,\varpi}(\beta_{2,n} + M_\sigma s(p)/p_1 \wedge p_2)$ *for some large fixed constants* $C_{1,\varpi}$ *and* $C_{2,\varpi}$*. Then we have for large* n*, with probability greater than* $1-\delta$,

$$\|\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\mathbf{\Gamma}} \le C\left\{\frac{p\beta_{1,n}^2 + p_2\beta_{2,n}^2}{p^{1/2}} + \frac{p_2 s^2(p)}{p^{1/2}(p_1\wedge p_2)^2} + M_\sigma s(p)\alpha_n^{1-q}\right\}, \tag{4.11}$$

$$\|\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\max} \le C\alpha_n, \tag{4.12}$$

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 \le C M_\sigma s(p)\alpha_n^{1-q}, \tag{4.13}$$

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} \le C\alpha_n, \tag{4.14}$$

where the relative Frobenius norm $\|\mathbf{U}\|_{\mathbf{\Gamma}}^2 = p^{-1}\|\mathbf{\Gamma}^{-1/2}\mathbf{U}\mathbf{\Gamma}^{-1/2}\|_F^2$*, and* $\alpha_n = \beta_{2,n} + M_\sigma s(p)/p_1 \wedge p_2$*. Furthermore, if the smallest eigenvalues of* $\widetilde{\mathbf{\Gamma}}$ *and* $\widehat{\mathbf{\Sigma}}$ *are positive, we have for large* n,

with probability greater than $1-\delta$,

$$\|\widetilde{\mathbf{\Gamma}}^{-1} - \mathbf{\Gamma}^{-1}\|_2 \leq CM_\sigma s(p)\alpha_n^{1-q}, \tag{4.15}$$

$$\|\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|_2 \leq CM_\sigma s(p)\alpha_n^{1-q}. \tag{4.16}$$

Remark 6. Theorem 4.2 shows the consistency of the SPOET in terms of the relative Frobenius norm. For example, when both p_1 and p_2 have the order of p, we have, with probability greater than $1-p^{-1}$,

$$\|\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\mathbf{\Gamma}} \leq C\left\{\frac{p^{1/2}\log p}{n^{a/2}} + M_\sigma s(p)\left(\sqrt{\frac{\log p}{n^{a/2}}} + M_\sigma\frac{s(p)}{p}\right)^{1-q}\right\}.$$

The SPOET estimator is consistent so long as $p = o(n^a)$.

Remark 7. The diagonal condition (4.10) may be too restrictive in analyzing volatilities. So when relaxing the sparse condition (4.10) to the following sparse condition

$$\max_{j\in\mathcal{L}}\sum_{i\in\mathcal{L}}|\Sigma_{ij}|^q(\Sigma_{ii}\Sigma_{jj})^{(1-q)/2} \leq M_\sigma s(p),$$

the convergence rates corresponding to the sparse volatility matrix $\mathbf{\Sigma}$ are changed. For example, we have the term $\sqrt{\log p/n^{ab/2}} + M_\sigma s(p)/p_1 \wedge p_2$ instead of α_n in the results of Theorem 4.2. When $p_1 \asymp p_2$, we have, with probability greater than $1-p^{-1}$,

$$\|\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\mathbf{\Gamma}} \leq C\left\{\frac{p^{1/2}\log p}{n^{a/2}} + M_\sigma s(p)\left(\sqrt{\frac{\log p}{n^{ab/2}}} + M_\sigma\frac{s(p)}{p}\right)^{1-q}\right\}.$$

5 Simulation study

5.1 Consistency of estimators

To check the finite sample performance of the proposed estimator, we conducted a simulation study. The true log-stock price follows a continuous-time r-factor model defined in (2.1) with

$\boldsymbol{\mu}(t) = 0$. Let $\boldsymbol{\sigma}(t)$ be the Cholesky decomposition of the instantaneous volatility process $\boldsymbol{\varsigma}(t) = (\varsigma_{ij}(t))_{1\leq i,j\leq p}$. The diagonal elements of $\boldsymbol{\varsigma}(t)$ follow four different processes such as geometric Ornstein-Uhlenbeck processes, the sum of two CIR processes (Barndorff-Nielsen, 2002; Cox et al., 1985), 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 obtain the sparse integrated volatility matrix $\boldsymbol{\Sigma}$, we generated the off-diagonal elements as follows:

$$\varsigma_{ij}(t) = \begin{cases} 0, & \text{if } i,j \in \mathcal{L} \\ \{\kappa(t)\}^{|i-j|}\sqrt{\varsigma_{ii}(t)\varsigma_{jj}(t)}, & \text{otherwise,} \end{cases}$$

where the process $\kappa(t)$ is

$$\kappa(t) = \frac{e^{\frac{1}{2}u(t)} - 1}{e^{\frac{1}{2}u(t)} + 1}, \quad du(t) = 0.03\{0.64 - u(t)\}dt + 0.118u(t)dW_{\kappa,t},$$
$$W_{\kappa,t} = \sqrt{0.96}W^0_{\kappa,t} - 0.2\sum_{i=1}^{p} W_{it}/\sqrt{p},$$

and $W^0_{\kappa,t}, \kappa = 1,\ldots,p$, are one-dimensional Brownian motions which are independent of the Brownian motions $\mathbf{W}^*_t$ and $\mathbf{W}_t$. The low-rank instantaneous volatility matrix $\boldsymbol{\varsigma}^f(t) = \boldsymbol{\vartheta}^\top(t)\boldsymbol{\vartheta}(t)$ is $\mathbf{H}^\top\{\boldsymbol{\vartheta}^f(t)\}^\top\boldsymbol{\vartheta}^f(t)\mathbf{H}$, where $\mathbf{H} = (H_{ij})_{1\leq i\leq r, 1\leq j\leq p} \in \mathbb{R}^{r\times p}$ and H_{ij} were generated from i.i.d. uniform distribution on $[-2, 2]$. $\boldsymbol{\vartheta}^f(t)$ was generated similarly to $\boldsymbol{\sigma}(t)$. For example, $\boldsymbol{\vartheta}^f(t)$ is a diagonal matrix, and its squared diagonal elements were generated from three different processes: geometric Ornstein-Uhlenbeck processes, the sum of two CIR processes, and the volatility process in Nelsons GARCH diffusion limit model.

We generated the noisy high-frequency data $Y_i(t_k)$ by adding a noise term $\epsilon_i(t_k)$ obtained from independent normal distribution with mean zero and standard deviation $0.1\sqrt{\boldsymbol{\Gamma}_{ii}}$. To generate the non-synchronized data, we randomly selected the non-synchronized observation time points from the synchronized observation time points $t_k = \frac{k}{n^{all}}, k = 1,\ldots,n^{all}-1$. For example, the number of observation time points for each asset is determined by $\lfloor \pi_i n^{all} \rfloor$, where the proportion $\pi_i \in (0, 1)$. For liquid assets, the proportion π_i was generated from

i.i.d. uniform distribution $(0.5, 1)$, while for illiquid assets, the proportion π_i was generated from i.i.d. uniform distribution $(\frac{5L}{\sqrt{n^{all}}}, \frac{10L}{\sqrt{n^{all}}})$, where the liquidity level L was varied from 0.25 to 2. Then we obtained the non-synchronized sample path by randomly sampling $\lfloor \pi_i n^{all} \rfloor$ observation time points from $\{t_1, t_2, \ldots, t_{n^{all}-1}\}$.

We fixed the proportion of liquid assets to be 0.5, that is, $p_1 = p/2$. Using the simulated noisy non-synchronized data $Y_i(t_{i,k}), i = 1, \ldots, p, k = 1, \ldots, n_i$, we calculated the PRVM, defined in Definition 2 with the weight function $g(x) = x \wedge (1 - x)$ and $K = \lfloor n^{1/2} \rfloor$. Then we applied the proposed SPOET procedure and POET procedure. The latter regularizes directly the PRVM estimator $\widehat{\mathbf{\Gamma}}$. For the thresholding step, we used the adaptive hard thresholding scheme and chose the optimal thresholding level for each method by minimizing the corresponding Frobenius norm of the difference between the estimate and true value. In the simulation study, we fixed $p = 200$, $r = 3$, and $n^{all} = 23400$ which equals to the number of seconds in one day's trading period. The simulation process was repeated 500 times. The average numbers of synchronized observations after applying the pairwise refresh time scheme for liquid-liquid, liquid-illiquid, illiquid-illiquid combinations are reported in Table 1.

Table 1: Average sample sizes after applying the refresh time scheme for liquid-liquid, liquid-illiquid, and illiquid-illiquid combinations.

L	liquid-liquid	liquid-illiquid	illiquid-illiquid
0.25	14300.82	286.27	186.75
0.5	14292.87	573.04	374.48
1	14305.63	1145.44	752.32
2	14295.88	2279.87	1517.39

Figure 1 depicts the average estimation errors of the SPOET and POET for estimating the low-rank volatility matrix $\mathbf{\Theta}$ against the liquidity level L and the numerical results are reported in Table 2. It can easily be seen that the SPOET outperforms the POET (relative efficient greater than one) except one case where $L = 2$ using spectral norm. In terms of the Frobenius norm, the SPOET gets more efficiency than the POET as the liquidity level

L decreases. In fact, when the liquidity level decreases, both SPOET and POET estimators have larger average errors. However, the SPOET has smaller increment of errors than the POET, as the SPOET does not use the illiquid-illiquid block data but infers volatility in this block from the low-rank structure. Thus, it is more robust to the liquidity level L. On the other hand, the performance in terms of the spectral norm is relatively stable over the liquidity level L. To check the effect of the projection of $\widetilde{\mathbf{\Theta}}_{12}$ onto the space spanned by $\widehat{\mathbf{\Theta}}_{11}$, we compare the SPOET with A-SPOET (alternative SPOET in (4.6)). The projected low-rank estimator $\widehat{\mathbf{\Theta}}_{12}$ shows better performance than $\widetilde{\mathbf{\Theta}}_{12}$. From this result, we can find that the projection onto the accurate estimator $\widehat{\mathbf{\Theta}}_{11}$ helps to improve the performance of estimating low-ranking volatility $\mathbf{\Theta}$.

Table 2: Average errors under Frobenius norm and spectral norm of the SPOET, A-SPOET, and POET for $\mathbf{\Theta}$ with $L = 0.25, 0.5, 1, 2$.

		$\mathbf{\Theta}$			$\mathbf{\Theta}_{11}$		$\mathbf{\Theta}_{22}$		$\mathbf{\Theta}_{12}$		
	L	SPOET	A-SPOET	POET	SPOET	POET	SPOET	POET	SPOET	A-SPOET	POET
Frobenius	0.25	437.18	519.76	525.59	119.79	177.72	293.91	343.83	211.41	289.62	251.19
	0.5	367.59	421.48	426.16	119.76	151.45	234.92	270.16	180.48	231.80	206.77
	1	313.33	347.68	353.97	119.80	133.95	188.54	216.86	155.02	188.03	173.47
	2	269.98	290.40	297.53	119.69	121.83	150.51	174.20	133.79	153.67	147.04
Spectral	0.25	297.96	329.07	326.38	74.98	99.58	219.82	230.02	143.57	181.05	155.61
	0.5	239.34	258.64	263.53	75.11	86.26	165.79	179.94	119.60	143.18	128.37
	1	196.57	207.73	218.08	75.02	78.01	126.47	143.33	100.32	115.00	107.74
	2	162.20	167.90	181.27	74.87	72.36	95.09	113.11	83.64	92.38	90.65

Table 3 reports the average estimation errors, measured by the Frobenius, spectral, relative Frobenius norms, of the SPOET, A-SPOET, POET, and PRVM estimators for $\mathbf{\Gamma}$, $\mathbf{\Gamma}_{11}$, $\mathbf{\Gamma}_{12}$, and $\mathbf{\Gamma}_{22}$. Figure 2 shows the average errors of estimates for the integrated volatility matrix $\mathbf{\Gamma}$ based on the SPOET, A-SPOET, POET, and PRVM procedures for different liquidity levels L. As what we expected, SPOET, A-SPOET, and POET usually show better performance than PRVM. Furthermore, SPOET has the smallest average errors among these four estimators.

Finally, we compare performances of estimating the sparse volatility matrix $\mathbf{\Sigma}$ and inverse matrices $\mathbf{\Gamma}^{-1}$ and $\mathbf{\Sigma}^{-1}$. We report average estimation errors in Table 4, using both

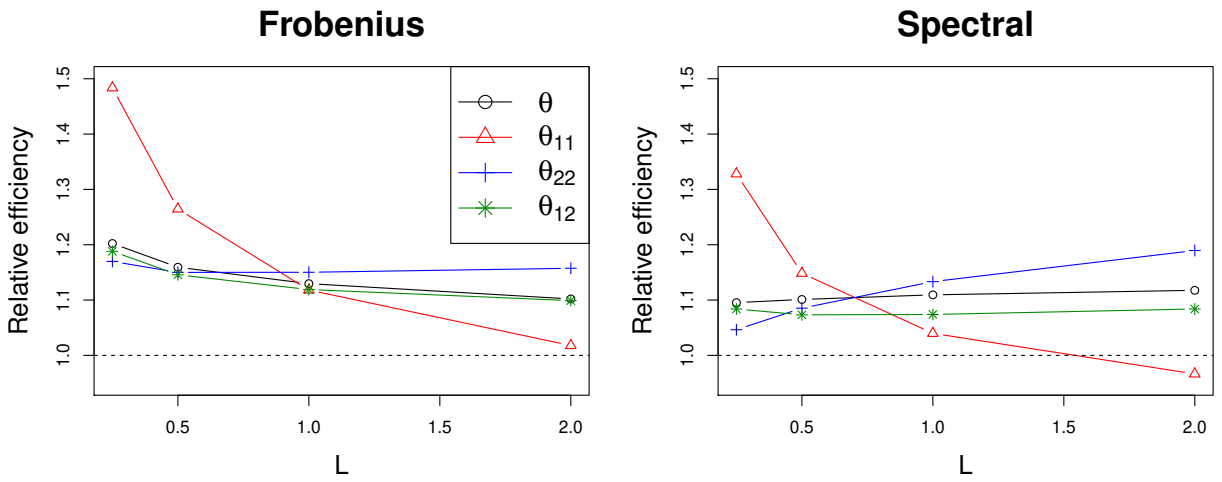

Figure 1: Relative efficiency of the SPOET with respect to the POET for estimating $\mathbf{\Theta}$ against the liquidity level L.

Table 3: The average estimaton errors of the SPOET, A-SPOET, POET, and PRVM for $\mathbf{\Gamma}$ using different matrix norms with $L = 0.25, 0.5, 1, 2$.

	Frobenius												
	$\mathbf{\Gamma}$				$\mathbf{\Gamma}_{11}$			$\mathbf{\Gamma}_{22}$			$\mathbf{\Gamma}_{12}$		
L	SPOET	A-SPOET	POET	PRVM	SPOET	POET	PRVM	SPOET	POET	PRVM	SPOET	POET	PRVM
0.25	441.9	523.6	527.9	790.1	128.1	182.8	162.3	297.5	344.7	478.5	211.4	251.2	429.4
0.5	372.5	425.8	429.6	671.3	128.1	158.3	162.3	238.3	271.7	404.1	180.5	206.8	361.2
1	318.5	352.3	358.2	572.5	128.1	142.1	162.3	191.7	218.5	341.5	155.1	173.5	303.9
2	275.4	295.5	302.4	488.2	128.0	130.9	162.3	153.4	175.9	285.8	133.8	147.1	255.2
	Spectral												
0.25	294.2	325.5	319.7	313.9	71.5	91.7	58.2	216.2	224.5	210.0	143.6	155.6	182.9
0.5	236.6	256.1	259.0	251.7	71.6	80.9	58.2	163.6	176.2	166.6	119.6	128.4	145.3
1	194.4	205.8	214.9	205.8	71.5	74.3	58.2	125.3	140.7	134.4	100.3	107.7	117.0
2	160.3	166.3	179.0	168.7	71.3	69.6	58.0	94.6	111.3	107.3	83.6	90.6	94.4
	Relative Frobenius												
0.25	1.23	1.65	1.60	2.94	0.56	0.86	0.91	1.13	1.44	2.35	-	-	-
0.5	1.03	1.31	1.26	2.49	0.55	0.73	0.91	0.88	1.09	1.98	-	-	-
1	0.88	1.06	1.04	2.13	0.55	0.65	0.91	0.70	0.86	1.67	-	-	-
2	0.76	0.88	0.87	1.82	0.56	0.60	0.91	0.55	0.67	1.40	-	-	-

Frobenius and spectral norms. Similar to the previous results, the SPOET usually shows better performance than other estimation procedures. However, when the liquidity level L is large ($L = 2$), the performances of the SPOET and POET procedures are similar. This is

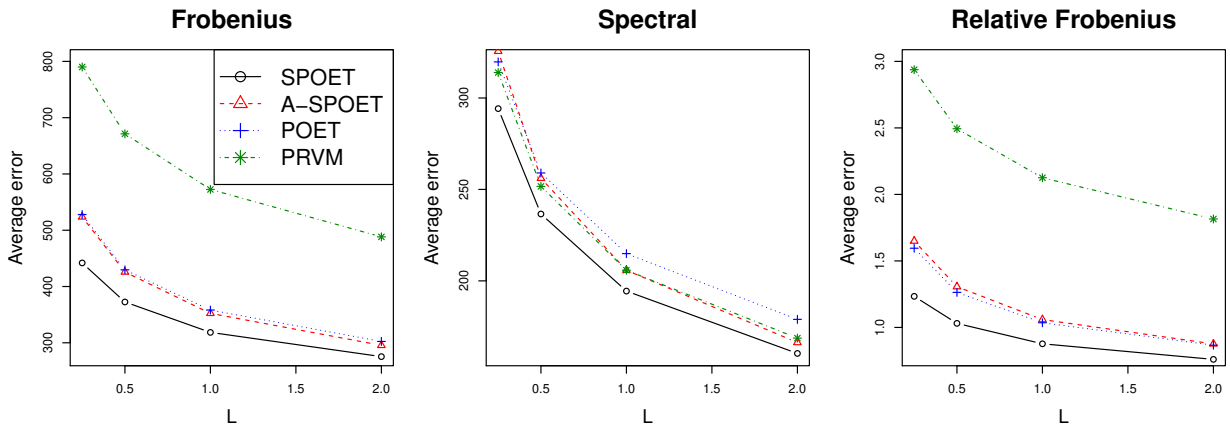

Figure 2: The average estimation errors of the SPOET, A-SPOET, POET, and PRVM for $\mathbf{\Gamma}$ using different matrix norms against the liquidity level L.

understandable: when the liquidity level is large, there is no big benefit from using accurate estimates to reconstruct the low-rank volatility matrix $\mathbf{\Theta}$.

Table 4: Average errors under Frobenius and spectral norms of the SPOET, POET, and PRVM for $\mathbf{\Sigma}$, $\mathbf{\Sigma}^{-1}$ and $\mathbf{\Gamma}^{-1}$ with $L = 0.25, 0.5, 1, 2$.

		$\mathbf{\Sigma}$		$\mathbf{\Sigma}^{-1}$		$\mathbf{\Gamma}^{-1}$		
	L	SPOET	POET	SPOET	POET	SPOET	POET	PRVM
Frobenius	0.25	82.379	107.408	0.786	5.321	1.362	10.074	58.239
	0.5	74.629	88.109	0.612	4.113	0.547	3.342	67.054
	1	69.349	75.566	0.577	1.279	0.507	0.807	82.767
	2	65.904	67.645	0.573	0.588	0.502	0.554	54.519
Spectral	0.25	17.098	23.652	0.495	4.932	1.074	9.713	56.479
	0.5	16.466	19.490	0.361	3.845	0.296	3.089	64.938
	1	16.416	17.154	0.349	1.039	0.280	0.566	80.336
	2	16.417	15.976	0.356	0.360	0.287	0.327	51.740

5.2 Portfolio risks

In this section, we further compared the SPOET, A-SPOET, POET, and PRVM for volatility matrix estimation using the portfolio risks as evaluation. Specifically, for each simulation setting, we generate 200 random portfolios $\mathbf{w} = (w_1, \ldots, w_p)^\top$ (approximately) uniformly

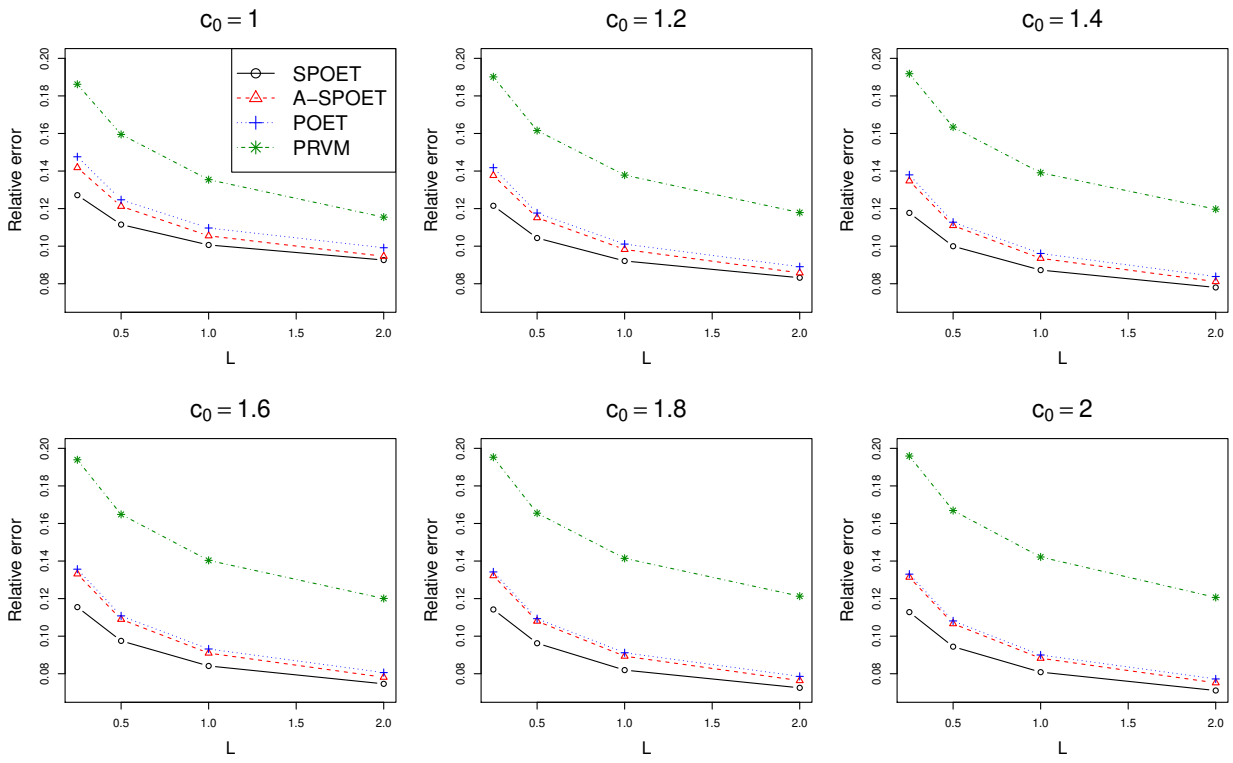

Figure 3: Average relative errors of estimators for the portfolio risk calculated using the SPOET, A-SPOET, POET, and PRVM estimators against the liquidity level L with the gross exposures $c_0 = 1, 1.2, 1.4, 1.6, 1.8, 2$, $p = 200$, and $r = 3$.

from the set $\{\mathbf{w} : \sum_{i=1}^{p} w_i = 1 \text{ and } \|\mathbf{w}\|_1 = c_0\}$, where c_0 is a given gross exposure. That is accomplished as follows. See Fan et al. (2015) for details and derivations. The number, k, of long positions is determined by a realization from binomial distribution $Bin(p, \frac{c_0+1}{2c_0})$. Then we generated independently $\{E_i\}_{i=1,\ldots,p}$ from standard exponential distributions. For the k long positions, the weight $w_i = (c_0 + 1)E_i/(2\sum_{j=1}^{k} E_j)$, $i = 1, \ldots, k$, and for the short positions, $w_i = -(c_0 - 1)E_i/(2\sum_{j=k+1}^{p} E_j)$, $i = k + 1, \ldots, p$. Finally, randomly permute those weights $\{w_i\}_{i=1}^{p}$.

For each of 500 simulated sample paths, we generated 200 testing portfolios, and so we have 100,000 portfolios in total for each estimation method. We varied the gross exposure c_0 from 1 to 2. For each portfolio, we calculated the relative error of estimated risk using the estimate $\widehat{\mathbf{\Gamma}}$ by $\frac{|\mathbf{w}^\top(\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma})\mathbf{w}|}{\mathbf{w}^\top\mathbf{\Gamma}\mathbf{w}}$, where $\widehat{\mathbf{\Gamma}}$ can be the SPOET, A-SPOET, POET, and PRVM. Then we computed the averages of $100,000$ errors as the performance measure for each method.

Figure 3 depicts the average relative errors of the portfolio risks calculated by the SPOET, A-SPOET, POET, and PRVM against the liquidity level L. We can find that the estimates based on the SPOET have the smallest error. As the liquidity level L increases, the difference between estimates based on the SPOET and POET estimators gets smaller. This is because when the liquidity level L is large, the illiquid part $\mathbf{\Gamma}_{22}$ is well estimated via POET procedure and so there is no huge benefit from using the structure of the low-rank matrix.

6 Empirical Applications

We collected intra-daily transaction prices of NYSE constituents from January to March in 2016 from the TAQ database in the Wharton Data Service (WRDS) system, 60 trading days in total. We excluded stocks which have less than 100 trading observations and chose the top 100 liquid stocks and the top 100 illiquid stocks as the candidates of our portfolio construction. We used the log-prices in seconds and exclude overnight returns to avoid dividend issuances and stock splits. To manage the non-synchronization problem, we used the pairwise refresh time. Average sample sizes for liquid-liquid, liquid-illiquid, and illiquid-illiquid blocks after applying the refresh time scheme are 6400, 615, and 313, respectively.

We calculated the SPOET, POET, and PRVM estimators for each trading day. For PRVM, we chose $g(x) = x \wedge (1-x)$ and $K = \lfloor n^{1/2} \rfloor$. For the thresholding step for the sparse volatility matrix $\mathbf{\Sigma}$, we used two different thresholding techniques for each of SPOET and POET that avoid the choice of thresholding parameters (Fan et al., 2016a): block diagonal, and block diagonal but using the diagonal part of estimated $\mathbf{\Sigma}_{22}$. We denoted the latter block diagonal threshold estimators by SPOET+Block and POET+Block. Blocks are determined using the Global Industry Classification Standard (GICS) (Fan et al., 2016a; Ait-Sahalia and Xiu, 2017). The idiosyncratic components for different blocks (sector) are set to zero and for the same block are untouched (Fan et al., 2016a). To determine the number, r, of factors, we calculated 60 integrated volatility matrices using the PRVM estimation procedure. Then we used the average eigenvalues from 60 PRVM estimators and draw the scree plot, which is shown in Figure 4. From Figure 4, we can see that the number of leading factors is around

5. In the empirical study, we chose $r = 1, 2, 3, 4, 5, 6$ for sensitivity analysis, though it is known that slight overestimate of the number of factors does no little harm to the portfolio choice(Fan et al., 2013).

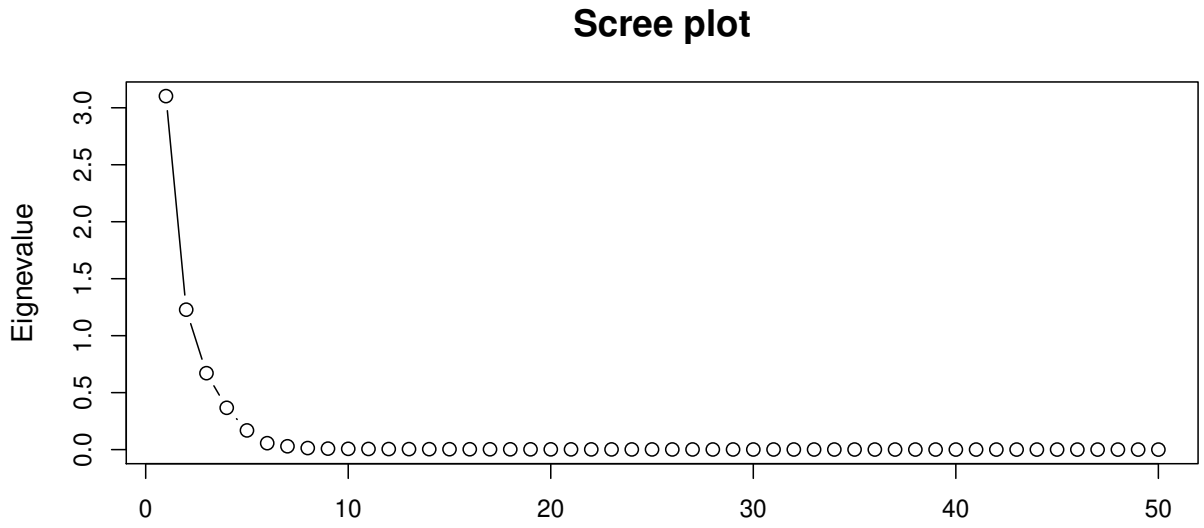

Figure 4: The scree plot of average eigenvalues of 60 PRVMs.

We examined the performance of the integrated volatility matrix estimators in a minimum variance portfolio allocation problem. We consider the following constrained minimum variance portfolio allocation problem:

$$\min_{\omega} \omega^\top \widehat{\mathbf{\Gamma}} \omega, \quad \text{subject to } \omega^\top \mathbf{J} = 1 \text{ and } \|\omega\|_1 = c_0,$$

where $\mathbf{J} = (1, \ldots, 1)^\top \in \mathbb{R}^p$, the gross exposure constraint c_0 was varied from 1 to 3, and $\widehat{\mathbf{\Gamma}}$ could be SPOET, POET, and PRVM. To make the estimates positive semi-definite, we projected the sparse volatility estimators for SPOET and POET, and PRVM estimator onto the positive semi-definite cone in the spectral norm. We constructed the portfolio at the beginning of each trading day and held it for one day. We calculated the standard deviation using the open-to-close log-returns of the portfolios, which is used to measure the portfolio risk.

Figure 5 depicts the out-of-sample risks of the portfolios constructed by SPOET, SPOET+Block,

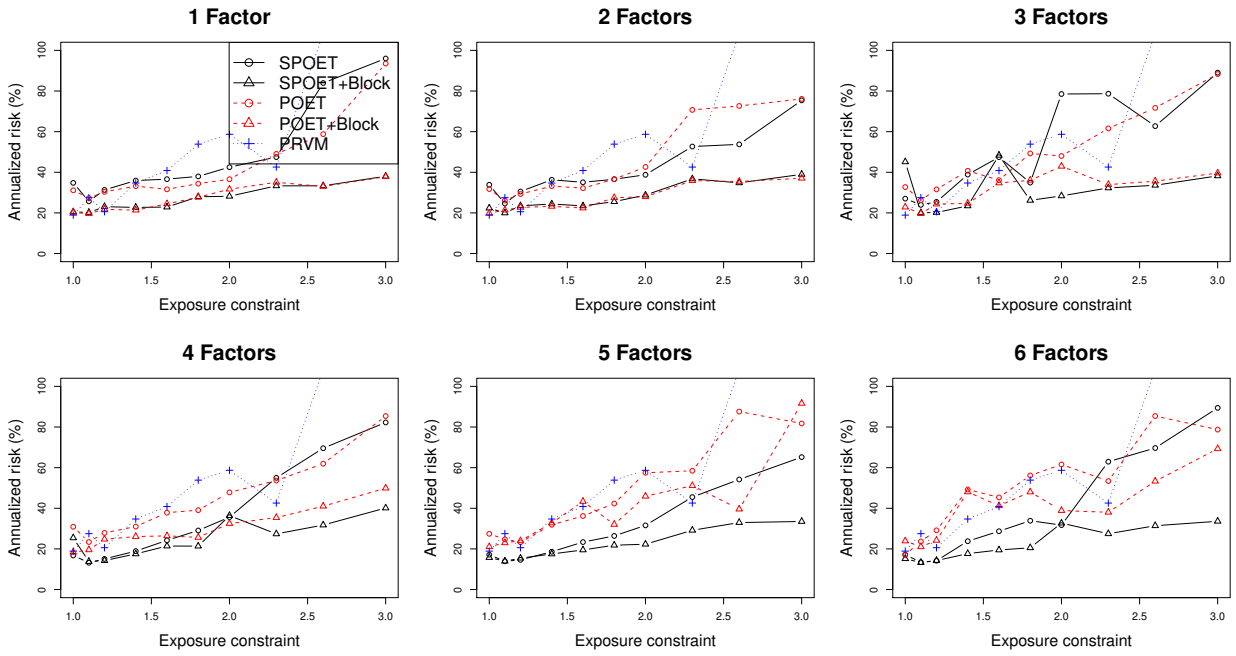

Figure 5: The out-of-sample risks of the optimal portfolios constructed by using the volatility matrix from SPOET, SPOET+Block, POET, POET+Block, and PRVM estimators with $r = 1, \ldots, 6$.

POET, POET+Block, and PRVM against the exposure constraint c_0. The minimum risks for portfolios constructed by using SPOET, SPOET+Block, POET, POET+Block, and PRVM over the c_0 are 13.2 %, 13.27%, 17.19 %, 18.34 %, and 18.91%, respectively. The SPOET estimation method reduces the minimum risks by 30%–43%. We can find that for the purpose of portfolio allocation, the SPOET and POET type estimators perform well and improve the performance of the PRVM. In addition, the PRVM estimator becomes unstable as the exposure constraint increases. When comparing thresholding schemes, the block thresholding scheme is generally better than taking $\mathbf{\Sigma}_{22}$ to be diagonal, which indicates that the block diagonal assumption is an appropriate assumption for stock returns. Meanwhile, when the number of factor is 5, the SPOET shows the stable results and performs better than others. The results suggest that the proposed SPOET procedure can help estimate the volatilities for illiquid assets.

7 Proofs

Proof of Theorem 4.1. Similar to the proof of Theorem 4.1 of Fan and Kim (2017), we can show

$$\|\widehat{\mathbf{\Theta}}_{11} - \mathbf{\Theta}_{11}\|_{\max} \leq C\left(M_\sigma \frac{s(p)}{p_1} + \beta_{1,n}\right).$$

Now consider $\|\widehat{\mathbf{\Theta}}_{12} - \mathbf{\Theta}_{12}\|_{\max}$. By Weyl's Theorem, we have

$$\begin{aligned}
\max_{1\leq k\leq r} |\widehat{\xi}_k - \xi_k| &\leq \|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Theta}_{12}\|_2 \\
&\leq \|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Gamma}_{12}\|_2 + \|\mathbf{\Gamma}_{12} - \mathbf{\Theta}_{12}\|_2 \\
&\leq \sqrt{p_1 p_2}\|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Gamma}_{12}\|_{\max} + M_\sigma s(p). \qquad (7.1)
\end{aligned}$$

By Theorem 1.1 in Fan et al. (2016b), we have

$$\begin{aligned}
&\max_{1\leq k\leq r} \|\widehat{\mathbf{v}}_k - sign(\langle\widehat{\mathbf{v}}_k, \mathbf{v}_k\rangle)\mathbf{v}_k\|_{\max} \\
&\leq C\frac{\sqrt{p_1p_2}\|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Gamma}_{12}\|_{\max} + M_\sigma\sqrt{p_1p_2}\max(s(p)/p_1, s(p)/p_2)}{D_\xi\sqrt{p_1}} \\
&\leq C\sqrt{p_2}\frac{\|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Gamma}_{12}\|_{\max} + M_\sigma\max(s(p)/p_1, s(p)/p_2)}{D_\xi} \qquad (7.2)
\end{aligned}$$

and

$$\begin{aligned}
&\max_{1\leq k\leq r} \|\widehat{\mathbf{u}}_k - sign(\langle\widehat{\mathbf{u}}_k, \mathbf{u}_k\rangle)\mathbf{u}_k\|_{\max} \\
&\leq C\frac{\sqrt{p_1p_2}\|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Gamma}_{12}\|_{\max} + M_\sigma\sqrt{p_1p_2}\max(s(p)/p_1, s(p)/p_2)}{D_\xi\sqrt{p_2}} \\
&\leq C\sqrt{p_1}\frac{\|\widehat{\mathbf{\Gamma}}_{12} - \mathbf{\Gamma}_{12}\|_{\max} + M_\sigma\max(s(p)/p_1, s(p)/p_2)}{D_\xi}. \qquad (7.3)
\end{aligned}$$

Then simple algebraic manipulations show

$$\begin{aligned}
&\|\widehat{\mathbf{v}}_k\widehat{\mathbf{u}}_k^\top - \mathbf{v}_k\mathbf{u}_k\|_{\max} \\
&\leq \|\widehat{\mathbf{v}}_k - \mathbf{v}_k\|_{\max}\|\widehat{\mathbf{u}}_k - \mathbf{u}_k\|_{\max} + \|\widehat{\mathbf{v}}_k - \mathbf{v}_k\|_{\max}\|\mathbf{u}_k\|_{\max} + \|\mathbf{v}_k\|_{\max}\|\widehat{\mathbf{u}}_k - \mathbf{u}_k\|_{\max} \\
&\leq \frac{C}{D_\xi}\left\{\beta_{2,n} + M_\sigma\max(s(p)/p_1, s(p)/p_2)\right\}. \qquad (7.4)
\end{aligned}$$

By (7.1) and (7.4), we have

$$
\begin{aligned}
&\|\widetilde{\mathbf{\Theta}}_{12} - \mathbf{\Theta}_{12}\|_{\max} \\
&= \| \sum_{k=1}^{r} \left(\widehat{\xi}_k \widehat{\mathbf{v}}_k \widehat{\mathbf{u}}_k^\top - \xi_k \mathbf{v}_k \mathbf{u}_k^\top \right) \|_{\max} \\
&\leq \sum_{k=1}^{r} \|(\widehat{\xi}_k - \xi_k)(\widehat{\mathbf{v}}_k \widehat{\mathbf{u}}_k^\top - \mathbf{v}_k \mathbf{u}_k^\top)\|_{\max} + \|(\widehat{\xi}_k - \xi_k)\mathbf{v}_k \mathbf{u}_k^\top\|_{\max} + \|\xi_k(\mathbf{v}_k \mathbf{u}_k^\top - \widehat{\mathbf{v}}_k \widehat{\mathbf{u}}_k^\top)\|_{\max} \\
&\leq C \left\{ \beta_{2,n} + M_\sigma \max(s(p)/p_1, s(p)/p_2) \right\} \\
&\leq C \left\{ \beta_{2,n} + M_\sigma \left(\frac{s(p)}{p_1} + \frac{s(p)}{p_2} \right) \right\}. \qquad (7.5)
\end{aligned}
$$

Simple algebraic manipulations show

$$
\begin{aligned}
&\|\widehat{\mathbf{\Theta}}_{12} - \mathbf{\Theta}_{12}\|_{\max} \\
&\leq \|\mathbf{Q}\mathbf{Q}^\top(\widetilde{\mathbf{\Theta}}_{12} - \mathbf{\Theta}_{12})\|_{\max} + \|(\widehat{\mathbf{Q}}\widehat{\mathbf{Q}}^\top - \mathbf{Q}\mathbf{Q}^\top)\mathbf{\Theta}_{12}\|_{\max} \\
&\qquad + \|(\widehat{\mathbf{Q}}\widehat{\mathbf{Q}}^\top - \mathbf{Q}\mathbf{Q}^\top)(\widetilde{\mathbf{\Theta}}_{12} - \mathbf{\Theta}_{12})\|_{\max} \\
&= (a) + (b) + (c).
\end{aligned}
$$

For (a), we have

$$
\begin{aligned}
(a) &\leq \|\mathbf{Q}\mathbf{Q}^\top\|_1 \|\widetilde{\mathbf{\Theta}}_{12} - \mathbf{\Theta}_{12}\|_{\max} \\
&\leq C \left\{ \beta_{2,n} + M_\sigma \left(\frac{s(p)}{p_1} + \frac{s(p)}{p_2} \right) \right\}, \qquad (7.6)
\end{aligned}
$$

where the last inequality is due to (7.5) and Assumption 2(e). For (b), we have

$$
\begin{aligned}
(b) &\leq p_1 \|\widehat{\mathbf{Q}}\widehat{\mathbf{Q}}^\top - \mathbf{Q}\mathbf{Q}^\top\|_{\max} \|\mathbf{\Theta}_{12}\|_{\max} \\
&\leq C \frac{p_1}{D_\lambda} \left\{ \beta_{1,n} + M_\sigma s(p)/p_1 \right\} \frac{\xi_1}{\sqrt{p_1 p_2}} \\
&\leq C \left\{ \beta_{1,n} + M_\sigma s(p)/p_1 \right\}, \qquad (7.7)
\end{aligned}
$$

where the second inequality can be derived similar to the proof of (7.4). Finally, for (c), similar to the proofs of (7.6) and (7.7), we can show

$$
(c) \leq p_1 \|\widehat{\mathbf{Q}}\widehat{\mathbf{Q}}^\top - \mathbf{Q}\mathbf{Q}^\top\|_{\max} \|\widetilde{\mathbf{\Theta}}_{12} - \mathbf{\Theta}_{12}\|_{\max}
$$

$$
\begin{aligned}
&\leq C\frac{p_1}{D_\lambda}\left\{\beta_{1,n}+M_\sigma s(p)/p_1\right\}\left\{\beta_{2,n}+M_\sigma\left(\frac{s(p)}{p_1}+\frac{s(p)}{p_2}\right)\right\}\\
&= o((a)+(b)). \qquad (7.8)
\end{aligned}
$$

Thus, from (7.6)–(7.8), we have

$$
\|\widehat{\boldsymbol{\Theta}}_{12}-\boldsymbol{\Theta}_{12}\|_{\max}\leq C\left\{\beta_{1,n}+\beta_{2,n}+M_\sigma\left(\frac{s(p)}{p_1}+\frac{s(p)}{p_2}\right)\right\}.
$$

Consider $\|\widehat{\boldsymbol{\Theta}}_{22}-\boldsymbol{\Theta}_{22}\|_{\max}$. It follows that

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Theta}}_{22}-\boldsymbol{\Theta}_{22}\|_{\max}\\
&\leq \|(\widehat{\boldsymbol{\Theta}}_{21}-\boldsymbol{\Theta}_{21})\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Lambda}}^{-1}\widehat{\mathbf{Q}}^\top\widehat{\boldsymbol{\Theta}}_{12}\|_{\max}+\|\boldsymbol{\Theta}_{21}\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Lambda}}^{-1}\widehat{\mathbf{Q}}^\top(\widehat{\boldsymbol{\Theta}}_{12}-\boldsymbol{\Theta}_{12})\|_{\max}\\
&\quad+\|\boldsymbol{\Theta}_{21}\{\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Lambda}}^{-1}\widehat{\mathbf{Q}}^\top-\mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^\top\}\boldsymbol{\Theta}_{12}\|_{\max}\\
&=(I)+(II)+(III).
\end{aligned}
$$

For (I), we have for large n,

$$
\begin{aligned}
(I) &\leq \|\underset{r}{\widehat{\boldsymbol{\Theta}}}_{21}-\underset{r}{\boldsymbol{\Theta}}_{21}\|_{\max}\|\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Lambda}}^{-1}\widehat{\mathbf{Q}}^\top\widehat{\boldsymbol{\Theta}}_{12}\|_1\\
&\leq \sum_{k'=1}\sum_{k=1}\frac{\widehat{\xi}_{k'}}{\widehat{\lambda}_k}\|\widehat{\mathbf{q}}_k\widehat{\mathbf{q}}_k^\top\widehat{\mathbf{v}}_{k'}\widehat{\mathbf{u}}_{k'}^\top\|_1\|\widehat{\boldsymbol{\Theta}}_{21}-\boldsymbol{\Theta}_{21}\|_{\max}\\
&\leq C\frac{\xi_1}{\lambda_r}\max_{1\leq k,k'\leq r}\|\widehat{\mathbf{q}}_k\widehat{\mathbf{u}}_{k'}^\top\|_1\|\widehat{\boldsymbol{\Theta}}_{21}-\boldsymbol{\Theta}_{21}\|_{\max}\\
&\leq C\frac{\xi_1}{\lambda_r}p_1^{1/2}\max_{1\leq k\leq r}\|\mathbf{u}_k\|_{\max}\|\widehat{\boldsymbol{\Theta}}_{21}-\boldsymbol{\Theta}_{21}\|_{\max}\\
&\leq C\|\widehat{\boldsymbol{\Theta}}_{21}-\boldsymbol{\Theta}_{21}\|_{\max}\\
&\leq C\left\{\beta_{2,n}+M_\sigma\left(\frac{s(p)}{p_1}+\frac{s(p)}{p_2}\right)\right\},
\end{aligned}
$$

where the third inequality is due to (7.1) and Proposition 7.1 (Fan and Kim, 2017) and the fourth inequality is from (7.3). Similarly, we can show

$$
(II)\leq C\left\{\beta_{2,n}+M_\sigma\left(\frac{s(p)}{p_1}+\frac{s(p)}{p_2}\right)\right\}.
$$

For (III), we have

$$\begin{aligned}
(III) &\leq \|\boldsymbol{\Theta}_{12}\|_1^2 \|\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Lambda}}^{-1}\widehat{\mathbf{Q}}^\top - \mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^\top\|_{\max} \\
&\leq C\xi_1^2 \max_{1\leq k\leq r} \|\mathbf{v}_k\mathbf{u}_k^\top\|_1^2 \|\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Lambda}}^{-1}\widehat{\mathbf{Q}}^\top - \mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^\top\|_{\max} \\
&\leq C\frac{\xi_1^2 p_1}{p_2} \|\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Lambda}}^{-1}\widehat{\mathbf{Q}}^\top - \mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^\top\|_{\max} \\
&\leq C\frac{\xi_1^2 p_1}{p_2} \sum_{i=1}^r \left(\|(\widehat{\lambda}_i^{-1} - \lambda_i^{-1})\mathbf{q}_i\mathbf{q}_i^\top\|_{\max} + \|\widehat{\lambda}_i^{-1}(\widehat{\mathbf{q}}_i\widehat{\mathbf{q}}_i^\top - \mathbf{q}_i\mathbf{q}_i^\top)\|_{\max} \right) \\
&\leq C\frac{\xi_1^2 p_1}{p_2} \left(\frac{p_1\beta_{1,n} + M_\sigma s(p)}{p_1\lambda_r^2} + \frac{\beta_{1,n} + M_\sigma s(p)/p_1}{D_\lambda \lambda_r} \right) \\
&\leq C\frac{\xi_1^2 p_1}{p_2} \frac{1}{\lambda_r D_\lambda} \left(\beta_{1,n} + M_\sigma \frac{s(p)}{p_1} \right) \\
&\leq C \left(\beta_{1,n} + M_\sigma \frac{s(p)}{p_1} \right),
\end{aligned}$$

where the fifth inequality can be derived similar to (7.1) and (7.4). ■

Proof of Theorem 4.2. (4.12) and (4.14) are immediately proved by Theorem 4.1, Assumption 2(b), and Assumption 3(a).

Consider (4.13). By Theorem 4.1, Assumption 2(b), and Assumption 3(a), we have

$$\begin{aligned}
\|\widetilde{\boldsymbol{\Sigma}}_{11} - \boldsymbol{\Sigma}_{11}\|_{\max} &\leq C \left\{ \beta_{1,n} + M_\sigma \frac{s(p)}{p_1} \right\}, \\
\|\widetilde{\boldsymbol{\Sigma}}_{12} - \boldsymbol{\Sigma}_{12}\|_{\max} &\leq C \left\{ \beta_{2,n} + M_\sigma \frac{s(p)}{p_1 \wedge p_2} \right\}.
\end{aligned} \tag{7.9}$$

Then we have for $j \in \mathcal{H}$,

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

$$
\begin{aligned}
&+\sum_{i=1}^{p}|\Sigma_{ij}|^q\varpi_{ij}^{1-q} \\
&\leq \frac{3}{2}2^q\sum_{i=1}^{p}|\Sigma_{ij}|^q\varpi_{ij}^{1-q}+\sum_{i=1}^{p}|\Sigma_{ij}|\,\mathbf{1}\left(|\Sigma_{ij}|\leq\frac{3}{2}\varpi_{ij}\right)+\sum_{i=1}^{p}|\Sigma_{ij}|^q\varpi_{ij}^{1-q} \\
&\leq C\sum_{i=1}^{p}|\Sigma_{ij}|^q\varpi_{ij}^{1-q} \\
&\leq C\left(\sum_{i\in\mathcal{H}}|\Sigma_{ij}|^q\varpi_{ij}^{1-q}+\sum_{i\in\mathcal{L}}|\Sigma_{ij}|^q\varpi_{ij}^{1-q}\right) \\
&\leq CM_\sigma\Bigg[s(p)\left\{\beta_{1,n}+M_\sigma\frac{s(p)}{p_1}\right\}^{1-q} \\
&\qquad +s(p)\left\{\beta_{1,n}+\beta_{2,n}+M_\sigma\left(\frac{s(p)}{p_1}+\frac{s(p)}{p_2}\right)\right\}^{1-q}\Bigg],
\end{aligned}
$$

and similarly, for $j\in\mathcal{L}$,

$$
\begin{aligned}
&\sum_{i=1}^{p}|\widehat{\Sigma}_{ij}-\Sigma_{ij}| \\
&\leq C\Bigg[M_\sigma s(p)\left\{\beta_{1,n}+\beta_{2,n}+M_\sigma\left(\frac{s(p)}{p_1}+\frac{s(p)}{p_2}\right)\right\}^{1-q} \\
&\qquad\qquad +\left\{\beta_{1,n}+\beta_{2,n}+M_\sigma\left(\frac{s(p)}{p_1}+\frac{s(p)}{p_2}\right)\right\}\Bigg].
\end{aligned}
$$

Thus,

$$
\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_2\leq CM_\sigma s(p)\alpha_n^{1-q}.
$$

Consider (4.11). Simple algebraic manipulation shows

$$
\|\widetilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_{\boldsymbol{\Gamma}}\leq\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_{\boldsymbol{\Gamma}}+\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\boldsymbol{\Gamma}}.
$$

For $\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_{\boldsymbol{\Gamma}}$, similar to the proofs of Theorem 4.2 (Kim et al., 2017), we have

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_{\boldsymbol{\Gamma}} \\
&\leq C\left\{\frac{1}{p^{1/2}\lambda_r(\boldsymbol{\Theta})}\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_F+\frac{\lambda_1(\widehat{\boldsymbol{\Theta}})}{p^{1/2}\lambda_r(\boldsymbol{\Theta})^2}\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_F^2+\frac{\lambda_1(\widehat{\boldsymbol{\Theta}})}{p^{1/2}\lambda_r(\boldsymbol{\Theta})^{3/2}}\|\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}\|_F\right\}
\end{aligned}
$$

$$\leq C\left\{\frac{1}{p^{1/2}\lambda_r(\mathbf{\Theta})}\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_F^2+\frac{1}{p^{1/2}\lambda_r(\mathbf{\Theta})^{1/2}}\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_F\right\}.$$

For $\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_{\mathbf{\Gamma}}$, we have

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_{\mathbf{\Gamma}} &\leq p^{-1/2}\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\|_2\|\mathbf{\Gamma}^{-1}\|_F\\ &\leq CM_\sigma s(p)\alpha_n^{1-q},\end{aligned}$$

where the last inequality is due to (4.13). Therefore,

$$\begin{aligned}\|\widetilde{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_{\mathbf{\Gamma}} &\leq C\left\{\frac{\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_F^2}{p^{1/2}\lambda_r(\mathbf{\Theta})}+\frac{\|\widehat{\mathbf{\Theta}}-\mathbf{\Theta}\|_F}{p^{1/2}\lambda_r(\mathbf{\Theta})^{1/2}}+M_\sigma s(p)\alpha_n^{1-q}\right\}\\ &\leq C\big[p^{-3/2}\left\{p_1^2\beta_{1,n}^2+p_1p_2\beta_{2,n}^2+p_2^2(\beta_{1,n}^2+\beta_{2,n}^2)\right\}+\frac{p_2s^2(p)}{p^{1/2}(p_1\wedge p_2)^2}+M_\sigma s(p)\alpha_n^{1-q}\big],\end{aligned}$$

where the last inequality is due to Theorem 4.1.

The statements (4.15)and (4.16) can be shown similar to the proofs of Theorem 4.1 (Fan and Kim, 2017). ■

Acknowlegement. The research of Jianqing Fan was supported in part by NSF grant DMS-1712591 and an Princeton engineering innovation fund. The research of Donggyu Kim was supported in part by KAIST Settlement/Research Subsidies for Newly-hired Faculty grant G04170049. The bulk of the work was conducted while Donggyu Kim was a post-doctoral fellow at Department of Operations Research and Financial Engineering, Princeton University.

References

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-

time process in the presence of market microstructure noise. *The review of financial studies*, 18(2):351–416.

Ait-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics, forthcoming*.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Barndorff-Nielsen, O. E. (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, 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(6):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *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.

Cai, T., Cai, T. T., and Zhang, A. (2016). Structured matrix completion with applications to genomic data integration. *Journal of the American Statistical Association*, 111(514):621–633.

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772.

Chamberlain, G. and Rothschild, M. (1982). Arbitrage, factor structure, and mean-variance analysis on large asset markets.

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(1):116–133.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 53:385–407.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.

Fan, J., Furger, A., and 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. and Kim, D. (2017). Robust volatility matrix estimation with factor-based diffusion model for high-frequency financial data. *Journal of the American Statistical Association, forthcoming*.

Fan, J., Li, Y., and Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, 107(497):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 (Statistical Methodology)*, 75(4):603–680.

Fan, J., Liao, Y., and Shi, X. (2015). Risks of large portfolios. *Journal of Econometrics*, 186(2):367–387.

Fan, J., Wang, W., and Zhong, Y. (2016b). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *arXiv preprint arXiv:1603.03516*.

Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, 102(480):1349–1362.

Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379.

Hayashi, T. and Yoshida, N. (2011). Nonsynchronous covariation process and limit theorems. *Stochastic processes and their applications*, 121(10):2416–2454.

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of financial econometrics*, 3(4):456–499.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276.

Kim, D., Liu, Y., and Wang, Y. (2017). Large volatility matrix estimation with factor-based diffusion model for high-frequency financial data. *Bernoulli, forthcoming*.

Kim, D. and Wang, Y. (2016). Sparse pca based on high-dimensional itô processes with measurement errors. *Journal of Multivariate Analysis*, 152:172–189.

Kim, D., Wang, Y., and Zou, J. (2016). Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Processes and their Applications*, 126:3527–3577.

Koltchinskii, V., Lounici, K., and Tsybakov, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, pages 2302–2329.

Kong, X.-B. (2017). On the systematic and idiosyncratic volatilities with large panel high-frequency data. *The Annals of Statistics, forthcoming*.

Malliavin, P. and Mancino, M. E. (2002). Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics*, 6(1):49–61.

Malliavin, P., Mancino, M. E., et al. (2009). A fourier transform method for nonparametric estimation of multivariate volatility. *The Annals of Statistics*, 37(4):1983–2010.

Mancino, M. E. and Sanfelici, S. (2008). Robustness of fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics & data analysis*, 52(6):2966–2989.

Park, S., Hong, S. Y., and Linton, O. (2016). Estimating the quadratic covariation matrix for asynchronously observed high frequency stock returns corrupted by additive measurement error. *Journal of Econometrics*, 191(2):325–347.

Tao, M., Wang, Y., Zhou, H. H., et al. (2013). Optimal sparse volatility matrix estimation for high-dimensional itô 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. and 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, 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(472):1394–1411.