Econometric Theory, **0**, 2024, 1–16.
doi:10.1017/S0266466624000240

LARGE GLOBAL VOLATILITY MATRIX ANALYSIS BASED ON OBSERVATION STRUCTURAL INFORMATION

Sung hoon Choi
University of Connecticut

Donggyu Kim
University of California, Riverside

In this article, we develop a novel large volatility matrix estimation procedure for analyzing global financial markets. Practitioners often use lower-frequency data, such as weekly or monthly returns, to address the issue of different trading hours in the international financial market. However, this approach can lead to inefficiency due to information loss. To mitigate this problem, our proposed method, called Structured Principal Orthogonal complEment Thresholding (S-POET), incorporates observation structural information for both global and national factor models. We establish the asymptotic properties of the S-POET estimator, and also demonstrate the drawbacks of conventional covariance matrix estimation procedures when using lower-frequency data. Finally, we apply the S-POET estimator to an out-of-sample portfolio allocation study using international stock market data.

1. INTRODUCTION

Factor analysis and principal component analysis (PCA) are commonly used for various applications, including macroeconomic variable forecasting and portfolio allocation optimization (Stock and Watson, 2002; Bai, 2003; Fan, Furger, and Xiu, 2016). Recent research has highlighted the importance of considering local factors in addition to global factors. These local factors have an impact on individuals in each local group and can be defined by regional, country, or industry level (Kose, Otrok, and Whiteman, 2003; Bekaert, Hodrick, and Zhang, 2009; Fama and French, 2012; Moench, Ng, and Potter, 2013). To account for different level factors, a multilevel factor structure has been developed (Bai and Wang, 2015; Ando and Bai, 2016; Han, 2021).

Several large volatility matrix estimation procedures have been developed based on latent factor models to account for the strong cross-sectional correlation in the stock market (Fan et al., 2016; Ait-Sahalia and Xiu, 2017; Fan, Liu, and Wang, 2018a; Fan and Kim, 2019; Jung, Kim, and Yu, 2022). For instance, under the single-level factor model, Fan, Liao, and Mincheva (2013) proposed a

Address correspondence to Sung Hoon Choi, Department of Economics, University of Connecticut, Storrs, CT, USA; e-mail: sung_hoon.choi@uconn.edu.

principal orthogonal component thresholding (POET) covariance matrix estimation procedure when the factors are unobservable. This method can consistently estimate unobservable factors using PCA and cross-sectionally correlated errors via thresholding with a large number of assets. Recently, to account for the latent local factor structure, Choi and Kim (2023) developed a Double-POET (D-POET) covariance matrix estimation procedure based on the multilevel factor structure. Specifically, D-POET is a two-step estimation procedure by applying PCA at each factor level based on the block local factor structure.

The analysis of international financial markets is crucial for constructing global benchmark indices, such as the MSCI World. When analyzing global financial market data, it is a common practice to use lower frequencies, such as 2-day average, weekly, monthly, or quarterly data, instead of daily returns (Chib, Nardari, and Shephard, 2006; Bekaert et al., 2009; Hou, Karolyi, and Kho, 2011; Fama and French, 2012; Ando and Bai, 2017). This is because international stock markets operate at different times, which results in returns being based on different information sets when measured on any given date over a short period, such as daily. Hence, practitioners often opt for using weekly or monthly returns to mitigate the impact of non-synchronized trading hours in international markets. However, using lower-frequency data can lead to a loss of information and less efficient estimators. On the other hand, Burns, Engle, and Mezrich (1998) synchronized international stock prices using the conditional expected value and compared the non-synchronized and synchronized volatility matrices. In the high-frequency financial econometrics literature, several papers considered asynchronous financial data for covariance estimations by the sample splitting method (Aït-Sahalia, Fan, and Xiu, 2010; Fan, Li, and Yu, 2012a; Hautsch, Kyj, and Oomen, 2012; Lunde, Shephard, and Sheppard, 2016; Dai, Lu, and Xiu, 2019; Fan and Kim, 2019; Sun and Xu, 2022). However, unlike a high-frequency model setup, non-synchronized low-frequency observations can cause not only inefficiency but also inconsistency because an observation time gap is fixed. Thus, it is important to study the non-synchronized trading hours in international markets based on low-frequency observations and develop an efficient and effective large global volatility matrix estimation procedure.

This article proposes a novel large volatility matrix estimation procedure that incorporates observation structural information with an entire set of observations in the global and national factor model. Specifically, we consider the international financial market and impose a latent multilevel factor structure to account for both global and national risk factors. To handle the non-synchronized trading hours in the international market, we assume that the correlation is stationary with respect to the relative trading hour difference. For example, if the proportion of the overlapped time goes to one, its corresponding correlation converges to the correlation of the synchronized time, which is the parameter of interest. This structure helps theoretically understand the non-synchronized trading hour problem in the international market, and we study the large global volatility matrix estimation problem under the proposed stationary global and national factor model. For example, national factor membership is assumed to be naturally known, and we further assume that countries within the same continent have the same information set. To estimate the global volatility matrix, we first apply D-POET in each continental group using daily returns, which have the finest information in our model setup. To capture the spillover effect between continents, we conduct a low-rank approximation to each continental pair using a lower frequency, which helps mitigate the effect of the non-synchronized trading hours. Finally, to accommodate the latent global factors, we employ PCA on the structurally fitted global factor components from previous procedures, which we call Structured-POET (S-POET). We derive the rates of convergence for S-POET under different matrix norms and discuss its benefits. The empirical study on portfolio allocation supports the theoretical findings.

The remainder of the article is organized as follows: In Section 2, we introduce the model and propose the S-POET estimation procedure. Section 3 provides an asymptotic analysis of the S-POET estimator. In Section 4, we conduct a simulation study to evaluate the finite sample performance of the proposed method. Section 5 applies the proposed method to a real data problem of portfolio allocation using global stock market data. In Section 6, we conclude the article. We provide a key proof in Section 7. All the remaining proofs and miscellaneous materials are presented in the Supplementary Material.

2. MODEL SETUP AND ESTIMATION PROCEDURE

We consider a global and national factor model (Choi and Kim, 2023):

$$y_{it} = b_i' G_t + \lambda_i^{l\prime} f_t^l + u_{it}, \text{ for } i = 1, \ldots, p, t = 1, \ldots, T, \text{ and } l = 1, \ldots, L, \tag{2.1}$$

where y_{it} is the ith log price belonging to country l at time t; G_t is a $k \times 1$ vector of latent global factors, and b_i is the global factor loadings; f_t^l is an $r_l \times 1$ vector of latent national factors that affect individuals belonging to country l, and λ_i^l is the corresponding national factor loadings; and u_{it} is an idiosyncratic error term, which is uncorrelated with G_t and f_t^l. Throughout the article, global and national factors are uncorrelated, while their factor loadings may not be orthogonal to each other. In addition, we assume that the numbers of factors, k and r_l, are fixed and the group membership of the national factors is known. Then, we can stack the observations and write the model (2.1) in a vector form as follows:

$$y_t = \mathbf{B} G_t + \mathbf{\Lambda} F_t + u_t, \tag{2.2}$$

where $y_t = (y_t^{1\prime}, \ldots, y_t^{L\prime})'$, where $y_t^l = (y_{(\sum_{j=0}^{l-1} p_j+1)t}, \ldots, y_{(\sum_{j=0}^{l} p_j)t})'$, p_l is the number of assets for country l, and $p_0 = 0$; the $p \times k$ matrix $\mathbf{B} = (b_1, \ldots, b_p)'$; the $p \times r$ block diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\mathbf{\Lambda}^1, \ldots, \mathbf{\Lambda}^L)$, where $\mathbf{\Lambda}^l = (\lambda_1^l, \ldots, \lambda_{p_l}^l)'$ is a $p_l \times r_l$ matrix of local factor loadings for each l such that $r = \sum_{l=1}^{L} r_l$; the $r \times 1$ vector $F_t = (f_t^{1\prime}, \ldots, f_t^{L\prime})'$; and $u_t = (u_{1t}, \ldots, u_{pt})'$.

In this article, we are interested in the $p \times p$ covariance matrix of y_t and its inverse matrix:

$$\mathbf{\Sigma} = \mathbf{B}\mathrm{cov}(G_t)\mathbf{B}' + \mathbf{\Lambda}\mathrm{cov}(F_t)\mathbf{\Lambda}' + \mathbf{\Sigma}_u := \mathbf{\Sigma}_g + \mathbf{\Sigma}_l + \mathbf{\Sigma}_u, \qquad \textbf{(2.3)}$$

where $\mathbf{\Sigma}_u = (\sigma_{u,ij})_{p\times p}$ is a sparse idiosyncratic covariance matrix of u_t such that, for some $q \in [0,1)$, $m_p = \max_{i\le p}\sum_{j\le p}|\sigma_{u,ij}|^q$ diverges slowly (Bickel and Levina, 2008). We note that the correlation matrix of y_t can be obtained by

$$\mathbf{R}_0 = (\rho_{0,ij})_{p\times p} = \mathbf{D}_0^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{D}_0^{-\frac{1}{2}}, \qquad \textbf{(2.4)}$$

where $\mathbf{D}_0$ is the diagonal matrix consisting of the diagonal elements of $\mathbf{\Sigma}$. Importantly, the correlation between stocks i and j, denoted by $\rho_{0,ij}$ in (2.4), can be realized through synchronized observations. However, in the context of the international stock market, each stock exchange operates its own trading hours. Hence, stocks traded on different exchanges have distinct observation time points, making it challenging to estimate $\rho_{0,ij}$ if stocks i and j are not in the same region. To address this issue, we introduce the following model structure. The ith observations in region s is $\{y_{i,t+\delta_s}\}_{t=1}^T$, and $\delta_s \in [0,1)$ is the market close time for $s \in \{1,\ldots,S\}$. We assume that $\{(y_{1,t+\delta_1},\ldots,y_{p,t+\delta_S})'\}_{t\ge1}$ is stationary. Denote the estimable correlation by $\rho_{h,ij}$ for assets i and j that are located in regions $s,q \in \{1,\ldots,S\}$, respectively, where the relative time difference $h = \frac{|\delta_s-\delta_q|}{d}$ and the window size of frequency $d = T^{1-\alpha}$ for $\alpha \in (0,1]$. Finally, we impose the following Lipschitz condition for $\rho_{h,ij}$:

$$|\rho_{h,ij} - \rho_{0,ij}| \le Ch^{\beta}, \qquad \textbf{(2.5)}$$

for some $\beta > 0$ and a positive constant C. This model setup provides a mathematical framework to understand a fraction of the global market. For example, from the proposed model setup perspective, when using daily return data, $\rho_{h,ij}$ does not converge to the synchronized correlation $\rho_{0,ij}$. In contrast, when using lower-frequency data, $\rho_{h,ij}$ converges to $\rho_{0,ij}$. Thus, in practice, researchers often use weekly or monthly returns instead of daily returns to mitigate the effect of different trading hours based on daily transaction prices (Chib et al., 2006; Bekaert et al., 2009; Hou et al., 2011; Fama and French, 2012; Ando and Bai, 2017). However, this causes inefficiency. We discuss this inefficiency theoretically in Section 3.

In this article, for simplicity, we assume that stocks in the same continent have the same observation time points; hence, regional membership is the continent. Naturally, regional membership is known. In addition, we assume that the number of regions, S, is fixed. Given the regional membership, we can stack the observations by country within each continent. Then, we define the “estimable” correlation matrix as follows:

$$\mathbf{R}_h = \begin{bmatrix} \mathbf{R}_{0,11} & \mathbf{R}_{h,12} & \cdots & \mathbf{R}_{h,1S} \\ \mathbf{R}_{h,21} & \mathbf{R}_{0,22} & \cdots & \mathbf{R}_{h,2S} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{h,S1} & \mathbf{R}_{h,S2} & \cdots & \mathbf{R}_{0,SS} \end{bmatrix}, \qquad \textbf{(2.6)}$$

where $\mathbf{R}_{0,ss} = (\rho_{0,ij})_{p_s \times p_s}$ and $\mathbf{R}_{h,sq} = (\rho_{h,ij})_{p_s \times p_q}$ for $s, q \in \{1, \ldots, S\}$. For simplicity, we use the subscript notation h, which is a function of i, j, and d. We note that, for $s \neq q$, $\mathbf{R}_{h,sq}$ represents the spillover effect between continents s and q, and we assume that its rank is k_{sq}^*. Moreover, without loss of generality, each rank is at most equal to the number of global factors (i.e., $k_{sq}^* \leq k$). We denote the corresponding covariance matrix by $\boldsymbol{\Sigma}_h = \mathbf{D}_0^{\frac{1}{2}} \mathbf{R}_h \mathbf{D}_0^{\frac{1}{2}}$.

Let $\boldsymbol{\Sigma}^s$ be the covariance matrix for continent s, which is a $p_s \times p_s$ diagonal block of $\boldsymbol{\Sigma}$. We then decompose $\boldsymbol{\Sigma}^s$ as follows:

$$\boldsymbol{\Sigma}^s = \boldsymbol{\Sigma}_g^s + \boldsymbol{\Sigma}_l^s + \boldsymbol{\Sigma}_u^s, \quad \text{for } s = 1, \ldots, S, \tag{2.7}$$

where $\boldsymbol{\Sigma}_g^s$ is the global factor component, $\boldsymbol{\Sigma}_l^s$ is the national factor component, and $\boldsymbol{\Sigma}_u^s$ is the idiosyncratic component. The equation (2.7) represents the multilevel factor-based covariance matrix. Thus, when we consider markets that have the same observation time, we can directly apply D-POET proposed by Choi and Kim (2023) with all possible observations to estimate the covariance matrix $\boldsymbol{\Sigma}^s$. However, when analyzing the global stock market, lower-frequency data are often used to avoid the issue of varying trading hours, which causes inefficiency. To handle this issue, we propose a S-POET procedure to estimate $\boldsymbol{\Sigma}$ as presented in Algorithm 3.1. S-POET efficiently estimates global and local factor components by utilizing all observations and considering the block structure of the local factors.

Remark 2.1. In high-frequency finance, the non-synchronization problem is usually observed, and synchronization schemes have been developed (Aït-Sahalia et al., 2010; Fan et al., 2012a; Hautsch et al., 2012). In a high-frequency setup, the observation time gaps go to zero, and the synchronization scheme tries to maximize the number of synchronized observations. In contrast, in a low-frequency setup, a time gap is fixed, which makes it hard to study the low-frequency non-synchronized observation problem. To handle this, we introduce the relative time difference that goes to zero. Unlike the high-frequency case, as the relative time shrinks to zero, the number of synchronized observations also decreases. Thus, we need to choose appropriate subsampling frequency, which is different from the high-frequency synchronization schemes. However, after we choose appropriate subsampling frequency, the synchronized low-frequency observation looks like the synchronized high-frequency data. Thus, as in Hautsch et al. (2012), we can synchronize the data for each block and then employ D-POET. This estimation procedure has the same convergence rate with D-POET using lower-frequency data because the slowest convergence rate among the blocks dominates. To overcome this, we propose S-POET that harnesses the multilevel factor structure. Specifically, the low-frequency non-synchronization is due to the regional difference that can account for the local factor in the multilevel factor structure. Thus, S-POET uses this observation structural information to account for the multilevel factor dynamics. In Section 3, we show a theoretical benefit of S-POET.

Algorithm 1 Structured-POET estimation procedure

1: For each continent s, we compute the Double-POET estimator (Choi and Kim, 2023) using T observations and denote it as $\widehat{\boldsymbol{\Sigma}}^{s,\mathcal{D}} \equiv \widehat{\boldsymbol{\Sigma}}_g^{s,\mathcal{D}} + \widehat{\boldsymbol{\Sigma}}_l^{s,\mathcal{D}} + \widehat{\boldsymbol{\Sigma}}_u^{s,\mathcal{D}}$. The specific procedure is described in Appendix S.2 in the online supplement. Let $\widetilde{\boldsymbol{\Sigma}}_g^{\mathcal{D}} = \operatorname{diag}(\widehat{\boldsymbol{\Sigma}}_g^{1,\mathcal{D}}, \ldots, \widehat{\boldsymbol{\Sigma}}_g^{S,\mathcal{D}})$, $\widetilde{\boldsymbol{\Sigma}}_l^{\mathcal{D}} = \operatorname{diag}(\widehat{\boldsymbol{\Sigma}}_l^{1,\mathcal{D}}, \ldots, \widehat{\boldsymbol{\Sigma}}_l^{S,\mathcal{D}})$, and $\widetilde{\boldsymbol{\Sigma}}_u^{\mathcal{D}} = \operatorname{diag}(\widehat{\boldsymbol{\Sigma}}_u^{1,\mathcal{D}}, \ldots, \widehat{\boldsymbol{\Sigma}}_u^{S,\mathcal{D}})$, where we denote $\operatorname{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_n)$ with the diagonal block entries as $\mathbf{A}_1, \ldots, \mathbf{A}_n$. Then, we construct a block diagonal matrix

$$\widetilde{\boldsymbol{\Sigma}}^{\mathcal{D}} = \operatorname{diag}(\widehat{\boldsymbol{\Sigma}}^{1,\mathcal{D}}, \ldots, \widehat{\boldsymbol{\Sigma}}^{S,\mathcal{D}}) \equiv \widetilde{\boldsymbol{\Sigma}}_g^{\mathcal{D}} + \widetilde{\boldsymbol{\Sigma}}_l^{\mathcal{D}} + \widetilde{\boldsymbol{\Sigma}}_u^{\mathcal{D}}. \tag{2.8}$$

2: Given a sample covariance matrix using d-day return data, $\widehat{\boldsymbol{\Sigma}}_h = T^{-\alpha} \sum_{t=1}^{T^\alpha} (y_t - \bar{y})(y_t - \bar{y})'$, we compute the sample correlation matrix $\widehat{\mathbf{R}}_h = \widehat{\mathbf{D}}_h^{-\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}_h \widehat{\mathbf{D}}_h^{-\frac{1}{2}}$, where $\widehat{\mathbf{D}}_h$ is the diagonal matrix consisting of the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_h$. We denote the sample correlation matrix $\widehat{\mathbf{R}}_h$ as the following block matrix form:

$$\widehat{\mathbf{R}}_h = (\widehat{\rho}_{h,ij})_{p \times p} = \begin{bmatrix} \widehat{\mathbf{R}}_{h,11} & \widehat{\mathbf{R}}_{h,12} & \cdots & \widehat{\mathbf{R}}_{h,1S} \\ \widehat{\mathbf{R}}_{h,21} & \widehat{\mathbf{R}}_{h,22} & \cdots & \widehat{\mathbf{R}}_{h,2S} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mathbf{R}}_{h,S1} & \widehat{\mathbf{R}}_{h,S2} & \cdots & \widehat{\mathbf{R}}_{h,SS} \end{bmatrix}.$$

For each (s,q)th off-diagonal partitioned block, we conduct the best rank-k_{sq}^* matrix approximation to $\widehat{\mathbf{R}}_{h,sq}$ such that $\widehat{\boldsymbol{\Theta}}_{sq} = \sum_{i=1}^{k_{sq}^*} \widehat{\xi}_i \widehat{u}_i \widehat{w}_i'$, where $\{\widehat{\xi}_i, \widehat{u}_i, \widehat{w}_i\}_{i=1}^{p_s \wedge p_q}$ are the ordered singular values, left-singular and right-singular vectors of $\widehat{\mathbf{R}}_{h,sq}$ in decreasing order. Then, we define

$$\widehat{\boldsymbol{\Theta}} = \begin{bmatrix} \mathbf{0} & \widehat{\boldsymbol{\Theta}}_{12} & \cdots & \widehat{\boldsymbol{\Theta}}_{1S} \\ \widehat{\boldsymbol{\Theta}}_{21} & \mathbf{0} & \cdots & \widehat{\boldsymbol{\Theta}}_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\boldsymbol{\Theta}}_{S1} & \widehat{\boldsymbol{\Theta}}_{S2} & \cdots & \mathbf{0} \end{bmatrix}.$$

3: Let $\widetilde{\delta}_1 \geq \widetilde{\delta}_2 \geq \cdots \geq \widetilde{\delta}_k$ be the k largest eigenvalues of $\widetilde{\boldsymbol{\Sigma}}_g = (\widetilde{\boldsymbol{\Sigma}}_g^{\mathcal{D}} + \widehat{\mathbf{D}}^{\frac{1}{2}} \widehat{\boldsymbol{\Theta}} \widehat{\mathbf{D}}^{\frac{1}{2}})$ and $\{\widetilde{v}_i\}_{i=1}^k$ be their corresponding eigenvectors, where $\widehat{\mathbf{D}}$ is the diagonal matrix consisting of the diagonal elements of (2.8). The final estimator of $\boldsymbol{\Sigma}$ is then defined as

$$\widehat{\boldsymbol{\Sigma}}^{\mathcal{S}} = \widetilde{\mathbf{V}}_g \widetilde{\boldsymbol{\Gamma}}_g \widetilde{\mathbf{V}}_g' + \widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}}, \tag{2.9}$$

where $\widetilde{\boldsymbol{\Gamma}}_g = \operatorname{diag}(\widetilde{\delta}_1, \ldots, \widetilde{\delta}_k)$, $\widetilde{\mathbf{V}}_g = (\widetilde{v}_1, \ldots, \widetilde{v}_k)$, $\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}} = \widetilde{\boldsymbol{\Sigma}}_l^{\mathcal{D}} + \widetilde{\boldsymbol{\Sigma}}_u^{\mathcal{D}}$, and $\widetilde{\boldsymbol{\Sigma}}_l^{\mathcal{D}}$ and $\widetilde{\boldsymbol{\Sigma}}_u^{\mathcal{D}}$ are defined in (2.8).

3. ASYMPTOTIC PROPERTIES

This section establishes the asymptotic properties of S-POET. Let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the minimum and maximum eigenvalues of matrix $\mathbf{A}$, respectively. We denote by $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|$, $\|\mathbf{A}\|_1$, and $\|\mathbf{A}\|_{\max}$ the Frobenius norm, operator norm, l_1-norm, and elementwise norm, which are defined, respectively, as $\|\mathbf{A}\|_F = \mathrm{tr}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_2 = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_1 = \max_j \sum_i |a_{ij}|$, and $\|\mathbf{A}\|_{\max} = \max_{i,j} |a_{ij}|$. When $\mathbf{A}$ is a vector, the maximum norm is denoted as $\|\mathbf{A}\|_\infty = \max_i |a_i|$, and both $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ are equal to the Euclidean norm. To investigate asymptotic behaviors, we require the following technical assumption.

Assumption 3.1.

(i) $\mathrm{cov}(G_t) = \mathbf{I}_k$, $\mathrm{cov}(f_t^l) = \mathbf{I}_{r_l}$, and $\mathbf{B}'\mathbf{B}$ and $\boldsymbol{\Lambda}^{l'}\boldsymbol{\Lambda}^l$ are diagonal matrices for $l \in \{1,\ldots,L\}$. In addition, G_t and f_t^l are uncorrelated with each other.

(ii) For some constants $c \in (0,1]$, $a_1 \in (\frac{3+2c}{5},1]$, and $a_2 \in (\frac{3}{5},1]$, all eigenvalues of $\mathbf{B}'\mathbf{B}/p^{a_1}$ and $\boldsymbol{\Lambda}^{l'}\boldsymbol{\Lambda}^l/p_l^{a_2}$ are strictly bigger than zero as $p, p_l \to \infty$, for $l \in \{1,\ldots,L\}$. In addition, $p_l \asymp p^c$, for each country l, and $a_1 \geq c a_2$. There is a constant $C > 0$ such that $\|\mathbf{B}\|_{\max} \leq C$ and $\|\boldsymbol{\Lambda}\|_{\max} \leq C$.

(iii) There exist constants $C_1, C_2 > 0$ such that $\lambda_{\min}(\boldsymbol{\Sigma}_u) > C_1$ and $\|\boldsymbol{\Sigma}_u\|_1 \leq C_2 m_p$.

(iv) Let $d = T^{1-\alpha}$ for $\alpha \in (0,1)$. The sample correlation matrix using d-day return data, $\widehat{\mathbf{R}}_h = \widehat{\mathbf{D}}_h^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_h\widehat{\mathbf{D}}_h^{-\frac{1}{2}}$, where $\widehat{\mathbf{D}}_h$ is the diagonal matrix consisting of the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_h = T^{-\alpha}\sum_{t=1}^{T^\alpha}(y_t - \bar{y})(y_t - \bar{y})'$, satisfies

$$\|\widehat{\mathbf{R}}_h - \mathbf{R}_h\|_{\max} = O_P(\sqrt{\log p/T^\alpha}).$$

(v) Denote $\boldsymbol{\Sigma} = (\Sigma_{ij})_{p\times p}$. The sample covariance matrix using T observations, $\widehat{\boldsymbol{\Sigma}} = T^{-1}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y})' = (\widehat{\Sigma}_{ij})_{p\times p}$, satisfies that, for $s \in \{1,\ldots,S\}$,

$$\max_{\{i,j\}\in s} |\widehat{\Sigma}_{ij} - \Sigma_{ij}| = O_P(\sqrt{\log p/T}).$$

Remark 3.1. Assumption 3.1(i) is the conventional normalization condition in the factor model literature. Assumption 3.1(ii) is known as the factor pervasiveness assumption, which is closely related to the incoherence structure (Fan, Wang, and Zhong, 2018b). This assumption can hold in macroeconomic and financial applications and is used for analyzing low-rank matrices (Chamberlain and Rothschild, 1983; Stock and Watson, 2002; Bai, 2003; Lam and Yao, 2012; Fan et al., 2013). Specifically, we allow the global and local factors to be weak by imposing technical conditions on a_1 and a_2. Intuitively, their lower bounds imply that both factors should have enough signals to satisfy the pervasive condition at different levels (see Choi and Kim, 2023). Assumptions 3.1(iv)–(v) provide a high-level sufficient condition for analyzing large matrices. The sample correlation matrix with d-day return data serves as the initial estimator for $\mathbf{R}_h$ in Assumption 3.1(iv). This condition is required to account for the spillover effect between continents. Here, the correlation matrix is considered to overcome the amplified scale issue of using the sample covariance matrix based on lower-frequency data (see Remark S.1 in the Supplementary Material). On the other hand, we can impose the element-wise convergence condition for each continent using the sample covariance matrix based on all observations (Assumption 3.1(iv)). These element-wise convergence rate conditions are easily satisfied under the sub-Gaussian condition and mixing time dependency (Vershynin, 2010; Fan et al., 2018a, 2018b). It can also be satisfied under heavy-tailed observations with bounded fourth moments (Fan et al., 2018a; Fan, Wang, and Zhu, 2021).

The following theorem provides the convergence rates for S-POET under various norms.

THEOREM 3.1. *Suppose that* $m_p = o(p^{c(5a_2-3)/2})$ *and Assumption* 3.1 *holds. Let* $\omega_T = p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1-\frac{3}{2}+c(\frac{5}{2}a_2-\frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}$. *If* $m_p\omega_T^{1-q} = o(1)$, *we have*

$$\|\widehat{\boldsymbol{\Sigma}}^{\mathcal{S}} - \boldsymbol{\Sigma}\|_{\max} = O_P\left(\omega_T + p^{5(1-a_1)}\Big(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{5a_1-4-c}}\right), \tag{3.1}$$

$$\begin{aligned}\|(\widehat{\boldsymbol{\Sigma}}^{\mathcal{S}})^{-1} - \boldsymbol{\Sigma}^{-1}\| \\ = O_P\left(m_p\omega_T^{1-q} + p^{\frac{c}{2}(1-a_2)}\omega_T + p^{\frac{11}{2}(1-a_1)}\Big(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{11}{2}a_1-\frac{9}{2}-c}}\right).\end{aligned} \tag{3.2}$$

In addition, if $a_1 > \frac{3}{4}$ *and* $a_2 > \frac{3}{4}$, *we have*

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}}^{\mathcal{S}} - \boldsymbol{\Sigma}\|_{\Sigma} = O_P\Big(& m_p\omega_T^{1-q} + p^{\frac{7}{2}(1-a_1)}\Big(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{7}{2}a_1-\frac{5}{2}-c}} \\ & + p^{\frac{21}{2}-10a_1}\Big(\frac{\log p}{T^{\alpha}} + \frac{1}{T^{2(1-\alpha)\beta}}\Big) + \frac{1}{p^{10a_1-\frac{17}{2}-2c}} + \frac{m_p^2}{p^{5ca_2-3c-\frac{1}{2}}}\Big),\end{aligned} \tag{3.3}$$

where the relative Frobenius norm is $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\Sigma} = p^{-1/2}\|\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2} - \mathbf{I}_p\|_F$.

Remark 3.2. ω_T is related to the estimation of latent local factors and idiosyncratic components using T observations. The additional terms $\sqrt{\log p/T^{\alpha}}$ and $1/T^{(1-\alpha)\beta}$ are the cost to handle the non-synchronized trading hours, when estimating latent global factors. In particular, the first term is coming from subsampled observations, $T^{\alpha} = T/d$, while the second term is the cost to estimate the synchronized correlation matrix $\mathbf{R}_0$. The optimal choice of α is $\alpha^* = \frac{2\beta}{1+2\beta}$, which simultaneously minimizes the convergence rates.

For simplicity, consider the case of strong global and local factors (i.e., $a_1 = 1$ and $a_2 = 1$), $q = 0$, and $m_p = O(1)$. Then, the proposed S-POET method yields

$$\begin{aligned}
&\|\widehat{\boldsymbol{\Sigma}}^{\mathcal{S}} - \boldsymbol{\Sigma}\|_{\Sigma} \\
&\quad = O_P\left(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}} + \frac{1}{p^{1-c}+p^{c}} + \sqrt{p}\Big(\frac{\log p}{T^{\alpha}} + \frac{1}{T^{2(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{3}{2}-2c}} + \frac{1}{p^{2c-\frac{1}{2}}}\right),
\end{aligned}$$

which can be convergent as long as $p = o(T^{\alpha})$ and $\frac{1}{4} < c < \frac{3}{4}$. Similar to the D-POET estimator (Choi and Kim, 2023), the upper and lower bounds of c are required to estimate both global and national factor components. To compare S-POET and D-POET, we derive the convergence rate of D-POET and demonstrate that the convergence rate of S-POET is faster than that of D-POET. Details can be found in Section S.3 of the Supplementary Material.

4. SIMULATION STUDY

In this section, simulations are carried out to examine the finite sample performance of S-POET. We generated non-synchronized observations as described in Section S.6 of the Supplementary Material. In this simulation study, we fixed the number of individuals $p = 500$. We set the number of continents $S = 2$ and the number of local groups $L = 20$ such that each continent group included 10 local groups (i.e., $p_l = 25$). Also, we chose the numbers of factors as $k = 3$ and $r = L \times r_l$, where $r_l = 2$ for each local group l. Then, we considered two cases: (i) increasing T from 100 to 600 in increments of 50 with the size of frequency $d \in \{1, 5\}$ (i.e., in-sample size is T/d) and (ii) increasing d from 1 to 10 with a fixed $T = 600$. For each case, 200 simulations were conducted.

For comparison, the sample covariance matrix (SamCov), POET, D-POET, and S-POET methods were employed to estimate $\boldsymbol{\Sigma}$. The average estimation errors were measured under $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\Sigma}$, $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max}$, and $\|(\widehat{\boldsymbol{\Sigma}})^{-1} - \boldsymbol{\Sigma}^{-1}\|$, where $\widehat{\boldsymbol{\Sigma}}$ is one of the covariance matrix estimators. We note that the lower-frequency data are obtained by summing the observations with d-day window. Therefore, for the SamCov estimator and the initial pilot estimator of POET and D-POET, we used $d^{-1}\widehat{\boldsymbol{\Sigma}}_h$ (see Section S.3 of the Supplementary Material). For each estimation, we determined the number of factors for D-POET and POET using the eigenvalue ratio methods suggested by Choi and Kim (2023) and Ahn and Horenstein (2013), with $k_{\max} = 10$ and $r_{l,\max} = 10$, respectively. For S-POET, we chose the number of ranks for each off-diagonal block by the largest singular value ratio. In addition, we employed the soft thresholding scheme for the idiosyncratic covariance matrix estimation.

Figures 1 and 2 depict the averages of estimation errors under different norms against T and d, respectively. From Figures 1 and 2, we find that S-POET has smaller estimation errors than the other methods. Specifically, in Figure 1, as T increases, the estimation errors of S-POET and estimators with $d = 5$ decrease, while the estimation errors of estimators with $d = 1$ do not decrease. This is

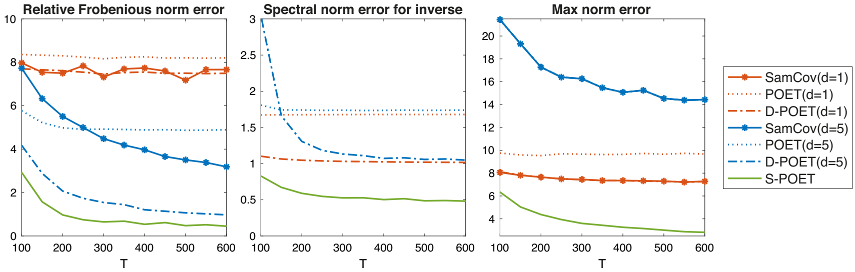

FIGURE 1. Averages of $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$, $\|(\widehat{\mathbf{\Sigma}})^{-1} - \mathbf{\Sigma}^{-1}\|$, and $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max}$ for SamCov, POET, D-POET, and S-POET against T with fixed $p = 500$ and $L = 20$. Lines that exceed the upper limits of the y-axis are excluded for the spectral norm error plot.

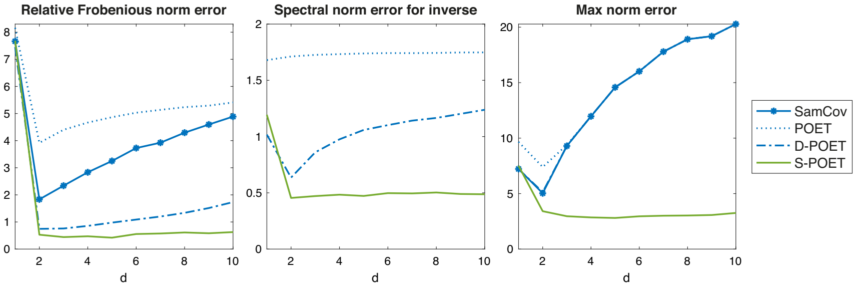

FIGURE 2. Averages of $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$, $\|(\widehat{\mathbf{\Sigma}})^{-1} - \mathbf{\Sigma}^{-1}\|$, and $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max}$ for SamCov, POET, D-POET, and S-POET against d with fixed $p = 500$, $T = 600$, and $L = 20$. Lines that exceed the upper limits of the y-axis are excluded for the spectral norm error plot.

because the estimators with $d = 1$ actually estimate $\mathbf{\Sigma}_h$ not $\mathbf{\Sigma}_0$, and when $d = 1$, $\mathbf{\Sigma}_h$ is not close to $\mathbf{\Sigma}_0$. When comparing the estimation procedures with $d = 5$, S-POET shows the best performance. This is because S-POET can accurately estimate the local factors and idiosyncratic components by utilizing whole observations, while other estimators utilize lower-frequency observations, which causes inefficiency. Figure 2 indicates that the estimation errors of all methods dramatically drop from $d = 1$ to $d = 2$, which is consistent with the results shown in Figure 1. S-POET shows stable results and has the minimum estimation errors when $d = 5$. In contrast, as the frequency size d increases, the estimation errors of SamCov, POET, and D-POET tend to increase again due to the smaller sample sizes. That is, the loss of information is severe only when using lower-frequency observations. From this result, we can conjecture that for a fixed T, the estimation error resulting from a small sample size with larger d is greater than the error resulting from the effect of observation time gaps. However, S-POET incorporates all available data to estimate the same regional covariance matrices, which helps enjoy the efficiency. It is worth noting that the estimation errors of POET seem constant as d increases. This is because POET does not estimate the local covariance matrix, which may dominate other estimation errors. The above results support the theoretical findings established in Section 3.

5. EMPIRICAL STUDY

We conducted a minimum variance portfolio allocation study using S-POET with global financial data. We obtained the daily transaction prices of international stock markets over 15 countries by the total market capitalization. The whole sample period is from 3 January 2017 to 30 December 2022. After excluding stocks with missing returns and no variation, we picked $1,500$ stocks based on the market cap for each country. In particular, we selected 500 firms for each continent and calculated both daily and weekly log-returns. The distribution of sample is presented in Table S.1 in the Supplementary Material.

We computed the S-POET, D-POET, POET, and SamCov estimators for each week. For S-POET, we used weekly or 2-day window returns to estimate the global factor component and daily returns to estimate the local factor and idiosyncratic components. We employed all daily, 2-day window, and weekly returns for the other procedures. For all POET-type procedures, we estimated the idiosyncratic volatility matrix using information of the 11 Global Industrial Classification Standard (GICS) sectors (Fan et al., 2016; Ait-Sahalia and Xiu, 2017). Specifically, we set the idiosyncratic components to zero for the different sectors, while maintaining them for the same sector. For a robustness check, we used different numbers of global factors, k, ranging from 1 to 5 for D-POET and from 1 to 20 for POET. For D-POET, we chose the number of local factors using the eigenvalue ratio method proposed by Ahn and Horenstein (2013) with $r_{l,\max} = 10$. In the S-POET procedure, for each off-diagonal partitioned block, we used the best rank-one approximation as suggested by the largest singular value gap method (see Section S.1 of the Supplementary Material).

We considered the following constrained minimum variance portfolio allocation problem (Fan, Zhang, and Yu, 2012b) to analyze the out-of-sample portfolio allocation performance:

$$\min_{\omega} \omega^T \widehat{\mathbf{\Sigma}} \omega, \text{ subject to } \omega^\top \mathbf{1} = 1, \ \|\omega\|_1 \le c,$$

where $\mathbf{1} = (1, \ldots, 1)^\top \in \mathbb{R}^p$, the gross exposure constraint c varies from 1 to 4, and $\widehat{\mathbf{\Sigma}}$ is one of the volatility matrix estimators obtained from S-POET, D-POET, POET, and SamCov. At the beginning of each week, we obtained optimal portfolios based on each estimator using the past 12 months' returns and held these portfolios for 1 week. We then computed the square root of the realized volatility using the weekly log-returns. Their averages were recorded for the out-of-sample risk. We examined six out-of-sample periods: 2018, 2019, 2020, 2021, 2022, and the full period from 2018 to 2022.

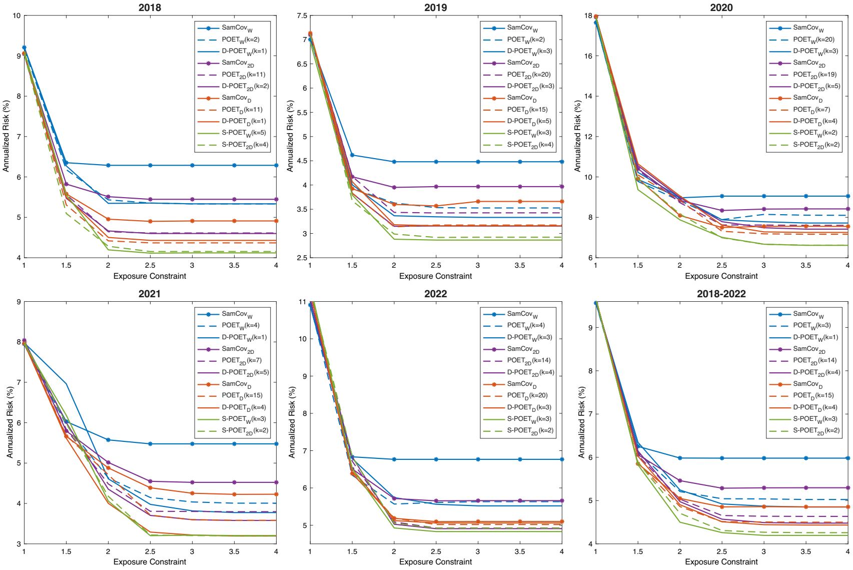

FIGURE 3. Out-of-sample risks of the optimal portfolios constructed by the SamCov, POET, Double-POET, and S-POET estimators for the global stock market.

Figure 3 illustrates the out-of-sample risks of the portfolios constructed by SamCov, POET, D-POET, and S-POET under varying exposure constraints. To draw readable plots, we presented the best performing results among the range of k for each estimator type and each period. We used subscripts to explicitly denote the frequency of the data used, with W (blue lines), 2D (purple lines), and D (red lines) representing weekly, 2-day window, and daily data, respectively. As shown in Figure 3, S-POET consistently outperforms the other estimators. Specifically, for all periods except 2021, S-POET$_W$ reduces the minimum risks by 4.7%–10.2% compared to the best estimator among the other methods. When comparing the estimation procedures with the same frequency observations, D-POET exhibits lower risks than POET and SamCov. Furthermore, D-POET, POET, and SamCov estimators using daily data tend to have lower risks than those using weekly data. This may be because, in practice, the impact of estimation inefficiency resulting from the smaller sample size could be greater than that resulting from the different observation time points. In addition, we also calculated the out-of-sample Sharpe ratio and averaged return in Table S.2 in the Supplementary Material. The results are based on selected estimators with minimum variances for the full period from 2018 to 2022. The results also indicate that S-POET$_W$ outperforms the other estimators. In summary, for portfolio allocation in the global stock market, S-POET, which incorporates both daily and weekly returns under the specific observation structure, outperforms POET and D-POET, which use only daily, 2-day window, or weekly returns. From this result, we can conjecture that the estimation accuracy for the national factor and idiosyncratic components can be improved using more frequent data (i.e., daily returns) for each continent group. In addition, for global factor estimation, using less frequent data (i.e., weekly returns) can manage the different trading hour problem.

6. CONCLUSION

In this article, we introduce a novel large global volatility matrix inference procedure. The proposed S-POET method leverages observation structural information from global financial markets based on latent global and national factor models. We establish the asymptotic properties of S-POET and demonstrate its efficiency in estimating a large global covariance matrix. In our empirical study, S-POET shows the best performance in terms of portfolio allocation. This is because S-POET accurately estimates the latent national factors and idiosyncratic components using daily returns. Additionally, using weekly returns to estimate the latent global factors can mitigate the effect of different trading hours across markets.

On the estimation procedure, we simplify the inter-continent correlation on the idiosyncratic term to reduce the estimation error because of the loss of information issue. However, it is interesting and important to handle both estimation error and model specification error, and thus, we leave this for a future study.

7. PROOFS

We establish (3.1) here and put the remaining proofs for (3.2) and (3.3) in Section S.4 of the Supplementary Material. Let $\{\bar{\delta}_i, \bar{v}_i\}_{i=1}^k$ and $\{\widetilde{\delta}_i, \widetilde{v}_i\}_{i=1}^k$ be the leading eigenvalues and eigenvectors of $\mathbf{B}\mathbf{B}'$ and $\widetilde{\mathbf{\Sigma}}_g$, respectively, where $\widetilde{\mathbf{\Sigma}}_g = (\widetilde{\mathbf{\Sigma}}_g^{\mathcal{D}} + \widehat{\mathbf{D}}^{\frac{1}{2}} \widehat{\mathbf{\Theta}} \widehat{\mathbf{D}}^{\frac{1}{2}})$.

LEMMA 7.1. *Under Assumption 3.1, for $i \le k$, we have*

$$|\widetilde{\delta}_i - \bar{\delta}_i| = O_P\left(p^{\frac{7}{2}-\frac{5}{2}a_1}\Big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{5}{2}a_1-\frac{5}{2}-c}}\right),$$

$$\|\widetilde{v}_i - \bar{v}_i\|_\infty = O_P\left(p^{5-\frac{11}{2}a_1}\Big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{11}{2}a_1-4-c}}\right).$$

Proof. Let $\widetilde{\mathbf{R}}_0 = (\tilde{\rho}_{0,ij})_{p\times p}$, where $\tilde{\rho}_{0,ij} = 0$ if $\{i,j\} \in s$, and $\tilde{\rho}_{0,ij} = \rho_{0,ij}$ if $i \in s$, $j \in q$, and $s \neq q$. By Assumption 3.1(iv), we have $\max_i |\widehat{\Sigma}_{ii} - \Sigma_{ii}| = O_P(\sqrt{\log p/T})$. Then, by Lemma S.3 in the Supplementary Material, we can easily obtain that

$$\|\widehat{\mathbf{D}}^{\frac{1}{2}} \widehat{\mathbf{\Theta}} \widehat{\mathbf{D}}^{\frac{1}{2}} - \mathbf{D}^{\frac{1}{2}} \widetilde{\mathbf{R}}_0 \mathbf{D}^{\frac{1}{2}}\|_{\max} = O_P\big(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta})\big). \qquad \textbf{(7.1)}$$

By using the fact that $\widetilde{\mathbf{\Sigma}}_g$ and $\mathbf{\Sigma}_g$ are low-rank matrices, (S.1) and (7.1), we have

$$
\begin{aligned}
&|\widetilde{\delta}_i - \bar{\delta}_i| \le \|\widetilde{\mathbf{\Sigma}}_g - \mathbf{\Sigma}_g\|_F \\
&\quad = O_P\left(\sqrt{\frac{p^2}{S}(p^{5(1-a_1)}\frac{\log p}{T} + \frac{1}{p^{5a_1-3-2c}}) + \frac{p^2(S-1)}{S}(p^{5(1-a_1)}(\frac{\log p}{T^{\alpha}} + \frac{1}{T^{2(1-\alpha)\beta}}))}\right) \\
&\quad = O_P\left(p\Big(p^{\frac{5}{2}(1-a_1)}(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}) + \frac{1}{p^{\frac{5}{2}a_1-\frac{3}{2}-c}}\Big)\right).
\end{aligned}
$$

By Theorem 1 of Fan et al. (2018b), (S.1), and (7.1), we have

$$
\|\widetilde{v}_i - \bar{v}_i\|_{\infty} \le Cp^{2(1-a_1)}\frac{\|\widetilde{\mathbf{\Sigma}}_g - \mathbf{\Sigma}_g\|_{\infty}}{p^{a_1}\sqrt{p}} = O_P\left(p^{5-\frac{11}{2}a_1}(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}) + \frac{1}{p^{\frac{11}{2}a_1-4-c}}\right).
$$

□

Proof of Theorem 3.1. Consider (3.1). Let $\mathbf{BB}' = \widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{V}}'$, where $\widetilde{\mathbf{\Gamma}} = \operatorname{diag}(\bar{\delta}_1, \ldots, \bar{\delta}_k)$ and their corresponding leading k eigenvectors $\widetilde{\mathbf{V}} = (\bar{v}_1, \ldots, \bar{v}_k)$. By Lemma 7.1, we have

$$
\begin{aligned}
&\|\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{BB}'\|_{\max} \\
&\quad \le \|\widetilde{\mathbf{V}}_g(\widetilde{\mathbf{\Gamma}}_g - \widetilde{\mathbf{\Gamma}})\widetilde{\mathbf{V}}_g'\|_{\max} + \|(\widetilde{\mathbf{V}}_g - \widetilde{\mathbf{V}})\widetilde{\mathbf{\Gamma}}(\widetilde{\mathbf{V}}_g - \widetilde{\mathbf{V}})'\|_{\max} + 2\|\widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}(\widetilde{\mathbf{V}}_g - \widetilde{\mathbf{V}})'\|_{\max} \\
&\quad = O(p^{-a_1}\|\widetilde{\mathbf{\Gamma}}_g - \widetilde{\mathbf{\Gamma}}\|_{\max} + \sqrt{p^{a_1}}\|\widetilde{\mathbf{V}}_g - \widetilde{\mathbf{V}}\|_{\max}) \\
&\quad = O_P\left(p^{5(1-a_1)}\Big(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{5a_1-4-c}}\right).
\end{aligned}
$$

In addition, by (S.2) and (S.3), we have $\|\widetilde{\mathbf{\Sigma}}_l^{\mathcal{D}} - \mathbf{\Lambda}\mathbf{\Lambda}'\|_{\max} = O_P(\omega_T)$ and $\|\widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u\|_{\max} = O_P(\omega_T)$. Therefore, we have

$$
\begin{aligned}
\|\widehat{\mathbf{\Sigma}}^{\mathcal{S}} - \mathbf{\Sigma}\|_{\max} &\le \|\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{BB}'\|_{\max} + \|\widetilde{\mathbf{\Sigma}}_l^{\mathcal{D}} - \mathbf{\Lambda}\mathbf{\Lambda}'\|_{\max} + \|\widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u\|_{\max} \\
&= O_P\left(p^{5(1-a_1)}\Big(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{5a_1-4-c}} + \omega_T\right).
\end{aligned}
$$

□

SUPPLEMENTARY MATERIAL

Choi, S. H., and Kim, D. (2024): Supplement to “Large Global Volatility Matrix Analysis Based on Observation Structural Information,” Econometric Theory Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466624000240.

REFERENCES

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81, 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, 1504–1517.

Ait-Sahalia, Y., & Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201, 384–399.

Ando, T., & Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31, 163–191.

Ando, T., & Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112, 1182–1198.

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

Bai, J., & Wang, P. (2015). Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, 33, 221–240.

Bekaert, G., Hodrick, R. J., & Zhang, X. (2009). International stock return comovements. *The Journal of Finance*, 64, 2591–2626.

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

Burns, P., Engle, R., & Mezrich, J. (1998). Correlations and volatilities of asynchronous data. *Journal of Derivatives*, 5, 7–18.

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

Chib, S., Nardari, F., & Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134, 341–371.

Choi, S. H., & Kim, D. (2023). Large volatility matrix analysis using global and national factor models. *Journal of Econometrics*, 235, 1917–1933.

Dai, C., Lu, K., & Xiu, D. (2019). Knowing factors or factor loadings, or neither? Evaluating estimators of large covariance matrices with noisy and asynchronous data. *Journal of Econometrics*, 208, 43–79.

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105, 457–472.

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

Fan, J., & Kim, D. (2019). Structured volatility matrix estimation for non-synchronized high-frequency financial data. *Journal of Econometrics*, 209, 61–78.

Fan, J., Li, Y., & Yu, K. (2012a). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, 107, 412–428.

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, 603–680.

Fan, J., Liu, H., & Wang, W. (2018a). Large covariance estimation through elliptical factor models. *The Annals of Statistics*, 46, 1383–1414.

Fan, J., Wang, W., & Zhong, Y. (2018b). An l_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18, 1–42.

Fan, J., Wang, W., & Zhu, Z. (2021). A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *The Annals of Statistics*, 49, 1239–1266.

Fan, J., Zhang, J., & Yu, K. (2012b). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107, 592–606.

Han, X. (2021). Shrinkage estimation of factor models with global and group-specific factors. *Journal of Business & Economic Statistics*, 39, 1–17.

Hautsch, N., Kyj, L. M., & Oomen, R. C. (2012). A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of Applied Econometrics*, 27, 625–645.

Hou, K., Karolyi, G. A., & Kho, B.-C. (2011). What factors drive global stock returns? *The Review of Financial Studies*, 24, 2527–2574.

Jung, K., Kim, D., & Yu, S. (2022). Next generation models for portfolio risk management: An approach using financial big data. *Journal of Risk and Insurance*, 89, 765–787.

Kose, M. A., Otrok, C., & Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors. *American Economic Review*, 93, 1216–1239.

Lam, C., & Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40, 694–726.

Lunde, A., Shephard, N., & Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, 34, 504–518.

Moench, E., Ng, S., & Potter, S. (2013). Dynamic hierarchical factor models. *The Review of Economics and Statistics*, 95, 1811–1817.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97, 1167–1179.

Sun, Y., & Xu, W. (2022). A factor-based estimation of integrated covariance matrix with noisy high-frequency data. *Journal of Business & Economic Statistics*, 40, 770–784.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint, arXiv:1011.3027*.