Large Global Volatility Matrix Analysis Based on Observation Structural Information

Sung Hoon Choi* Donggyu Kim†

September, 2023

Abstract

In this paper, 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 (Structured-POET), incorporates observation structural information for both global and national factor models. We establish the asymptotic properties of the Structured-POET estimator, and also demonstrate the drawbacks of conventional covariance matrix estimation procedures when using lower-frequency data. Finally, we apply the Structured-POET estimator to an out-of-sample portfolio allocation study using international stock market data.

Key words: High-dimensionality, international financial market, low-rank matrix, multi-level factor model, POET, sparsity.

*Department of Economics, University of Connecticut, Storrs, CT 06269, USA. E-mail: sung_hoon.choi@uconn.edu.

†College of Business, Korea Advanced Institute of Science and Technology (KAIST), Seoul, Republic of Korea. Email: donggyukim@kaist.ac.kr.

1 Introduction

Factor analysis and principal component analysis (PCA) are commonly used for various applications, including macroeconomic variable forecasting and portfolio allocation optimization (Bai, 2003; Bernanke et al., 2005; Fan et al., 2016; Stock and Watson, 2002). 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 (Bekaert et al., 2009; Fama and French, 2012; Kose et al., 2003; Moench et al., 2013). To account for different level factors, a multi-level factor structure has been developed. See Ando and Bai (2015, 2016); Bai and Wang (2015); Choi et al. (2018); Han (2021) for more information.

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 (Ait-Sahalia and Xiu, 2017; Fan et al., 2016, 2018a; Fan and Kim, 2019; Jung et al., 2022; Wang and Fan, 2017). For instance, under the single-level factor model, Fan et al. (2013) proposed a 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 covariance matrix estimation procedure based on the multi-level factor structure. Specifically, Double-POET is a two-step estimation procedure that estimates global and national factors separately 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 two-day average, weekly, monthly, or quarterly data, instead of daily returns (Ando and Bai, 2017; Bekaert et al., 2009; Chib et al., 2006; Choi and Kim, 2023; Fama and French, 2012; Hou et al., 2011). 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. Choi and Kim (2023) also used weekly returns to analyze the large global volatility matrix. However, using lower-frequency data can lead to a loss of information and less efficient estimators. Thus, it is important to study the non-synchronized trading hours in international markets and develop an efficient large global volatility matrix estimation procedure.

This paper 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 multi-level 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 the Double-POET procedure 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 Principal Orthogonal complEment Thresholding (Structured-POET). We derive the rates of convergence for the Structured-POET estimator under different matrix norms and discuss its benefits compared to the Double-POET procedure. For example, the Double-POET estimator may not be consistent using daily returns, while the Double-POET estimator with lower-frequency returns has a slower convergence rate than the Structured-POET estimator. An empirical study on portfolio allocation also supports the theoretical findings and shows the benefit of the proposed Structured-POET estimator.

The remainder of the paper is organized as follows. In Section 2, we introduce the model and propose the Structured-POET estimation procedure. Section 3 provides an asymptotic analysis of the Structured-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. Finally, the conclusion is provided in Section 6. All proofs are presented in the online supplementary file.

2 Model Setup and Estimation Procedure

Throughout this paper, let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the minimum and maximum eigenvalues of matrix $\mathbf{A}$, respectively. In addition, we denote by $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|_2$ (or $\|\mathbf{A}\|$ for short), $\|\mathbf{A}\|_1$, $\|\mathbf{A}\|_\infty$, and $\|\mathbf{A}\|_{\max}$ the Frobenius norm, operator norm, l_1-norm, l_∞-norm and elementwise norm, which are defined, respectively, as $\|\mathbf{A}\|_F = \mathrm{tr}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_2 = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_1 = \max_j \sum_i |a_{ij}|$, $\|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$, 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. We denote $\mathrm{diag}(\mathbf{A}_1, \dots, \mathbf{A}_n)$ with the diagonal block entries as $\mathbf{A}_1, \dots, \mathbf{A}_n$.

2.1 Multi-Level Factor Model

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 observed data for the ith individual belonging to country l at time t; G_t is a $k \times 1$ vector of unobserved latent global factors, and b_i is the global factor loadings; f_t^l is an $r_l \times 1$ vector of unobserved 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 paper, global and national factors are uncorrelated, while their factor loadings may not be orthogonal to each other. In addition, we assume that the group membership of the national factors is known and the numbers of factors, k and r_l, are fixed.

Due to the known national factor membership, 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 $y_t^l = (y_{p_1+\cdots+p_{l-1}+1t}, \ldots, y_{p_1+\cdots+p_l t})'$ and the number of individuals p_l within country l; the $p \times k$ matrix $\mathbf{B} = (b_1, \ldots, b_p)'$; the $p \times r$ block diagonal matrix $\mathbf{\Lambda} = \text{diag}(\Lambda^1, \ldots, \Lambda^L)$, where $\Lambda^l = (\lambda_1^l, \ldots, \lambda_{p_l}^l)'$ 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 $F_t = (f_t^{1\prime}, \ldots, f_t^{l\prime})'$, where f_t^l is an $r_l \times 1$ vector of local factors; and $u_t = (u_{1t}, \ldots, u_{pt})'$.

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

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

where $\mathbf{\Sigma}_u$ is a sparse idiosyncratic covariance matrix of u_t. In particular, we measure the sparsity level of $\mathbf{\Sigma}_u = (\sigma_{u,ij})_{p\times p}$ as follows (Bickel and Levina, 2008; Cai and Liu, 2011; Rothman et al., 2009): for some $q \in [0,1)$,

$$m_p = \max_{i\leq p} \sum_{j\leq p} |\sigma_{u,ij}|^q, \tag{2.4}$$

which diverges slowly, such as $\log p$. This implies that most pairs are weakly cross-sectionally correlated in the idiosyncratic error component. Decomposition (2.3) is a multi-level factor-based covariance matrix. Under the pervasive assumption, there are distinguished eigenvalues among the global factor components, the local factor components, and the idiosyncratic error components. Hence, we can analyze the model by the presence of distinguished eigenvalues at different levels (see Section 2.2). We note that the correlation matrix of y_t can be obtained by

$$\mathbf{R}_0 = (\rho_{0,ij})_{p\times p} = D_0^{-\frac{1}{2}} \mathbf{\Sigma} D_0^{-\frac{1}{2}}, \tag{2.5}$$

where 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 (2.5), 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.

We represent the ith observations in region s as $\{y_{i,t+\delta_s}\}_{t=1}^T$, and $\delta_s \in [0,1)$ is the market close time for $s \in \{1,\ldots,S\}$. We assume that $\{(y_{1,t+\delta_1},\ldots,y_{p,t+\delta_S})'\}_{t\geq 1}$ is stationary. We then 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 $\rho_{h,ij}$:

$$|\rho_{h,ij} - \rho_{0,ij}| \leq Ch^\beta, \tag{2.6}$$

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 we use daily return data (i.e., $d = 1$), $\rho_{h,ij}$ 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 (Ando and Bai, 2017; Bekaert et al., 2009; Chib et al., 2006; Fama and French, 2012; Hou et al., 2011). However, this causes inefficiency. We discuss this inefficiency theoretically in Section 3.

In this paper, 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:

$$\mathbf{R}_h = \begin{bmatrix} R_{0,11} & R_{h,12} & \cdots & R_{h,1S} \\ R_{h,21} & R_{0,22} & \cdots & R_{h,2S} \\ \vdots & \vdots & \ddots & \vdots \\ R_{h,S1} & R_{h,S2} & \cdots & R_{0,SS} \end{bmatrix}, \tag{2.7}$$

where $R_{0,ss} = (\rho_{0,ij})_{p_s \times p_s}$ and $R_{h,sq} = (\rho_{h,ij})_{p_s \times p_q}$ for $s, q \in \{1, \dots, S\}$. For simplicity, we use the subscript notation h, which is a function of i, j, and d. We note that, for $s \neq q$, $R_{h,sq}$ 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 $\mathbf{\Sigma}_h = D_0^{\frac{1}{2}} \mathbf{R}_h D_0^{\frac{1}{2}}$.

Let $\mathbf{\Sigma}^s$ be the covariance matrix for continent s, which is a $p_s \times p_s$ diagonal block of $\mathbf{\Sigma}$.

We then decompose $\mathbf{\Sigma}^s$ as follows:

$$\mathbf{\Sigma}^s = \mathbf{\Sigma}_g^s + \mathbf{\Sigma}_l^s + \mathbf{\Sigma}_u^s, \qquad \text{for } s = 1, \ldots, S, \tag{2.8}$$

where $\mathbf{\Sigma}_g^s$ is the global factor component, $\mathbf{\Sigma}_l^s$ is the national factor component, and $\mathbf{\Sigma}_u^s$ is the idiosyncratic component. The equation (2.8) represents the multi-level factor-based covariance matrix. Thus, when we consider markets that have the same observation time, we can directly apply the Double-POET procedure proposed by Choi and Kim (2023) with all possible observations (i.e., $d = 1$) to estimate the covariance matrix $\mathbf{\Sigma}^s$. Specifically, we first employ PCA to estimate $\mathbf{\Sigma}_g^s$ using the leading principal components based on the sample covariance matrix. Then, we apply PCA on each diagonal block of the remainder terms after removing the latent global factor components. Finally, we apply an adaptive thresholding method to the remaining idiosyncratic components. The specific algorithm is described in the online supplement. However, due to the problem of different trading hours, when analyzing the global stock market, lower-frequency data is often used, which causes inefficiency. To handle this issue, we propose a novel procedure for estimating large global covariance matrix $\mathbf{\Sigma}$ in the following subsection, which incorporates the structural information with an entire set of observations.

2.2 Structured-POET Procedure

Following Choi and Kim (2023), we assume the canonical conditions that $\text{cov}(G_t) = \mathbf{I}_k$, $\text{cov}(f_t^l) = \mathbf{I}_{r_l}$, and $\mathbf{B}'\mathbf{B}$ and $\Lambda^{l\prime}\Lambda^l$ are diagonal matrices for $l \in \{1, \ldots, L\}$. We note that G_t and f_t^l are uncorrelated with each other. Let $a_1, a_2 \in (0, 1]$ be the strengths of global and local factors, respectively. We then impose the following pervasiveness conditions: for each l, the eigenvalues of $p^{-a_1}\mathbf{B}'\mathbf{B}$ and $p_l^{-a_2}\Lambda^{l\prime}\Lambda^l$ are distinct and bounded away from zero. This condition implies that the first k eigenvalues of $\mathbf{B}\text{cov}(G_t)\mathbf{B}'$ diverge at rate $O(p^{a_1})$, while the first r eigenvalues of $\mathbf{\Lambda}\text{cov}(F_t)\mathbf{\Lambda}'$ diverge at rate $O(p^{ca_2})$, where $a_1 > ca_2$. For $c \in (0, 1]$, we note that $p^c \asymp p_l$ for each country l. Also, all eigenvalues of $\mathbf{\Sigma}_u$ are bounded.

To incorporate the structure of the global financial market discussed in Section 2.1, we propose a Structured-POET procedure to estimate $\mathbf{\Sigma}$ as follows:

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

$$\widetilde{\mathbf{\Sigma}}^{\mathcal{D}} = \mathrm{diag}(\widehat{\mathbf{\Sigma}}^{1,\mathcal{D}}, \dots, \widehat{\mathbf{\Sigma}}^{S,\mathcal{D}}) \equiv \widetilde{\mathbf{\Sigma}}_g^{\mathcal{D}} + \widetilde{\mathbf{\Sigma}}_l^{\mathcal{D}} + \widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}}. \tag{2.9}$$

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

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

   For each (s, q)th off-diagonal partitioned block, we conduct the best rank-k_{sq}^* matrix approximation to $\widehat{R}_{h,sq}$ such that $\widehat{\Theta}_{sq} = \sum_{i=1}^{k_{sq}^*} \widehat{\xi}_i \widehat{u}_i \widehat{w}_i'$, 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 $\widehat{R}_{sq}$ in decreasing order. Then, we define

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

3. Let $\widetilde{\delta}_1 \geq \widetilde{\delta}_2 \geq \cdots \geq \widetilde{\delta}_k$ be the k largest eigenvalues of $\widetilde{\mathbf{\Sigma}}_g = (\widetilde{\mathbf{\Sigma}}_g^{\mathcal{D}} + \hat{D}^{\frac{1}{2}}\widehat{\mathbf{\Theta}}\hat{D}^{\frac{1}{2}})$ and $\{\widetilde{v}_i\}_{i=1}^k$ be their corresponding eigenvectors, where $\hat{D}$ is the diagonal matrix consisting of the diagonal elements of (2.9). The final estimator of $\mathbf{\Sigma}$ is then defined as

$$\widehat{\mathbf{\Sigma}}^{\mathcal{S}} = \widetilde{\mathbf{V}}_g \widetilde{\mathbf{\Gamma}}_g \widetilde{\mathbf{V}}_g' + \widehat{\mathbf{\Sigma}}_E^{\mathcal{S}}, \tag{2.10}$$

where $\widetilde{\mathbf{\Gamma}}_g = \mathrm{diag}(\widetilde{\delta}_1, \ldots, \widetilde{\delta}_k)$, $\widetilde{\mathbf{V}}_g = (\widetilde{v}_1, \ldots, \widetilde{v}_k)$, $\widehat{\mathbf{\Sigma}}_E^{\mathcal{S}} = \widetilde{\mathbf{\Sigma}}_l^{\mathcal{D}} + \widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}}$, and $\widetilde{\mathbf{\Sigma}}_l^{\mathcal{D}}$ and $\widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}}$ are defined in (2.9).

Remark 2.1. To implement Structured-POET, we need to determine the rank k_{sq}^* and the number of factors, which are unknown in practice. We note that each (s,q)th off-diagonal partitioned block $R_{h,sq}$ in (2.7) is a low-rank matrix, and each rank is less than or equal to the number of global factors (i.e., $k_{sq}^* \leq k$). Thus, to determine the rank and number of global factors, we can use the data-driven methods proposed by Ahn and Horenstein (2013); Bai and Ng (2002); Onatski (2010). For example, the rank k_{sq}^* can be determined by finding the largest singular value gap such that $\max_{i \leq \bar{k}_{sq}}(\widehat{\xi}_i - \widehat{\xi}_{i+1})$, where $\bar{k}_{sq} = \min\{p_s, p_q\}$. On the other hand, to consistently estimate k, we employ the modified version of the eigenvalue ratio method, introduced by Choi and Kim (2023), based on $\widehat{\mathbf{\Sigma}}_h$.

In summary, for given continent and country memberships, we apply Double-POET to each continental block, which incorporates all observations to estimate the national factor and idiosyncratic components. We then conduct a low-rank approximation to each off-diagonal block using lower-frequency observations. Finally, we perform PCA on the combined global factor components obtained from the previous procedures. This procedure is called

Structured-POET. Structured-POET efficiently estimates global and local factor components by utilizing all observations and considering the block structure of the local factors. In contrast, the Double-POET or POET estimator may not be efficient due to the loss of the observation structural information. In Section 3.1, we discuss the theoretical inefficiency of Double-POET. Also, the numerical study in Sections 4 and 5 shows that Structured-POET outperforms Double-POET and POET.

3 Asymptotic Properties

This section establishes the asymptotic properties of the proposed Structured-POET estimator. To investigate asymptotic behaviors, we require the following technical assumption.

Assumption 3.1.

(i) *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* $\Lambda^{l\prime}\Lambda^{l}/p_l^{a_2}$ *are strictly bigger than zero as* $p, p_l \to \infty$, *for* $l \in \{1, \dots, L\}$. *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* $\|\mathbf{\Lambda}\|_{\max} \leq C$.

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

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

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

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

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

Remark 3.1. Assumption 3.1(i) is known as the factor pervasiveness assumption, which is closely related to the incoherence structure (Fan et al., 2018b). This assumption can hold in macroeconomic and financial applications and is used for analyzing low-rank matrices (Bai, 2003; Chamberlain and Rothschild, 1983; Fan et al., 2013, 2016; Lam and Yao, 2012; Stock and Watson, 2002). Specifically, in the context of a multi-level factor model, global factors have a broad impact on most individuals, while national factors only affect individuals within each national group. This illustrates the pervasive condition at different levels in an intuitive manner. Assumptions 3.1(iii)-(iv) 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(iii). 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 3.3). 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 (Fan et al., 2018a,b; Vershynin, 2010; Wang and Fan, 2017). It can also be satisfied under heavy-tailed observations with bounded fourth moments (Fan et al., 2017, 2018a,b, 2021).

In addition to the max norm and spectral norm, the relative Frobenius norm (Stein and James, 1961) is considered as a measure of large matrix estimation errors:

$$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\Sigma} = p^{-1/2} \|\boldsymbol{\Sigma}^{-1/2} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1/2} - \mathbf{I}_p\|_F.$$

Note that the factor $p^{-1/2}$ plays the role of normalization, i.e., $\|\boldsymbol{\Sigma}\|_{\Sigma} = 1$. Under this relative Frobenius norm, in the approximate single-level factor model, the conventional POET estimator is consistent if $p = o(T^2)$, while the sample covariance does not converge if $p > T$ (Fan et al., 2013). In the multi-level factor model, the Double-POET estimator is consistent as long as $p = o(T^2)$ and $\frac{1}{4} < c < \frac{3}{4}$, while the POET estimator does not converge if $c > \frac{1}{2}$ or $p^\kappa > T$ with $\kappa = \min\{\frac{1}{2}, 2c\}$ (Choi and Kim, 2023). In Section 3.1, we compare the convergence rates of Double-POET and Structured-POET.

We obtain the following convergence rates for Structured-POET and its inverse 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*

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

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

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

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

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 sub-sampled 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. This implies that the choice of frequency window d is important in practice. In the numerical study, we use the weekly data, that is, $d = 5$.

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 Structured-POET method yields

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

which can be convergent as long as $p = o(T^{\alpha})$ and $\frac{1}{4} < c < \frac{3}{4}$. Similar to the Double-POET estimator (Choi and Kim, 2023), the upper and lower bounds of c are required to estimate both global and national factor components.

3.1 Double-POET Using Lower-Frequency Data

In this subsection, we compare the Double-POET method with the proposed Structured-POET method.

To capture the global factor, local factor, and idiosyncratic components, we can apply the Double-POET method. However, when considering international stocks, practitioners commonly use lower-frequency data to minimize the impact of different observation time points. Let $\widehat{\boldsymbol{\Sigma}}_h = T^{-\alpha} \sum_{t=1}^{T^{\alpha}} (y_t - \bar{y})(y_t - \bar{y})'$ be the sample covariance matrix using d-day return data. Then, $d^{-1}\widehat{\boldsymbol{\Sigma}}_h$ is used for the initial pilot estimator for covariance matrix $\boldsymbol{\Sigma}$, since $\widehat{\boldsymbol{\Sigma}}_h$ is the amplified estimator by d, which slowly grows (see Remark 3.3). Let $\widehat{\boldsymbol{\Gamma}} = \mathrm{diag}(\widehat{\delta}_1, \ldots, \widehat{\delta}_k)$ and $\widehat{\mathbf{V}} = (\widehat{v}_1, \ldots, \widehat{v}_k)$ be the leading eigenvalues and their corresponding eigenvectors of $d^{-1}\widehat{\boldsymbol{\Sigma}}_h$. Next, let $\widehat{\boldsymbol{\Sigma}}_E^l$ be the lth $p_l \times p_l$ diagonal block of $\widehat{\boldsymbol{\Sigma}}_E = d^{-1}\widehat{\boldsymbol{\Sigma}}_h - \widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}'$. Let $\widehat{\Psi}^l = \mathrm{diag}(\widehat{\kappa}_1^l, \ldots, \widehat{\kappa}_{r_l}^l)$ and $\widehat{\Phi}^l = (\widehat{\eta}_1^l, \ldots, \widehat{\eta}_{r_l}^l)$ be the leading eigenvalues and their corresponding eigenvectors of $\widehat{\boldsymbol{\Sigma}}_E^l$. Let $\widehat{\boldsymbol{\Psi}} = \mathrm{diag}(\widehat{\Psi}^1, \ldots, \widehat{\Psi}^L)$, $\widehat{\boldsymbol{\Phi}} = \mathrm{diag}(\widehat{\Phi}^1, \ldots, \widehat{\Phi}^L)$, and $\widehat{\boldsymbol{\Sigma}}_u = d^{-1}\widehat{\boldsymbol{\Sigma}}_h - \widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' - \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}'$. Then, the Double-POET estimator is defined as follows:

$$\widehat{\boldsymbol{\Sigma}}^{\mathcal{D}} = \widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' + \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' + \widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}},$$

where $\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}}$ is the thresholded error covariance matrix estimator based on $\widehat{\boldsymbol{\Sigma}}_u = (\widehat{\sigma}_{u,ij})_{p \times p}$

(Bickel and Levina, 2008; Fan et al., 2013):

$$\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}} = (\widehat{\sigma}_{u,ij}^{\mathcal{D}})_{p\times p}, \qquad \widehat{\sigma}_{u,ij}^{\mathcal{D}} = \begin{cases} \widehat{\sigma}_{u,ij}, & i = j \\ s_{ij}(\widehat{\sigma}_{u,ij}) I(|\widehat{\sigma}_{u,ij}| \geq \tau_{ij}), & i \neq j \end{cases},$$

where an entry-dependent threshold $\tau_{ij} = \tau(\widehat{\sigma}_{u,ii}\widehat{\sigma}_{u,jj})^{1/2}$ and $s_{ij}(\cdot)$ is a generalized thresholding function such as hard thresholding ($s_{ij}(x) = x$), soft thresholding ($s_{ij}(x) = \mathrm{sgn}(x)(|x| - \tau_{ij})$, where $\mathrm{sgn}(\cdot)$ is the sign function) and the adaptive lasso (see Rothman et al., 2009). The thresholding constant is determined by $\tau \asymp \omega_{T^\alpha}$, where ω_{T^α} is defined in Theorem 3.2.

Assumption 3.2. *Let $d = T^{1-\alpha}$ for $\alpha \in (0,1)$. The sample covariance matrix using d-day return data, $\widehat{\boldsymbol{\Sigma}}_h = T^{-\alpha}\sum_{t=1}^{T^\alpha}(y_t - \bar{y})(y_t - \bar{y})'$, satisfies*

$$\|d^{-1}\widehat{\boldsymbol{\Sigma}}_h - \boldsymbol{\Sigma}_h\|_{\max} = O_P(\sqrt{\log p/T^\alpha}).$$

Remark 3.3. Assumption 3.2 is a necessary condition for the analysis of large covariance matrix inference. This element-wise convergence condition is similar to Assumption 3.1(iii) in Choi and Kim (2023). However, to match the scale of $\boldsymbol{\Sigma}_h$, the sample covariance matrix using d-day return data, $\widehat{\boldsymbol{\Sigma}}_h$, needs to be divided by d. To illustrate this point, consider the case when $p = 1$ and a collection of T i.i.d. random variables, $\{y_1, \dots, y_T\}$, where y_t is a log-return defined as $y_t = \log x_t - \log x_{t-1}$ and x_t is the asset price at time t. Assume that y_t has a mean of zero and a variance of σ^2. We can obtain lower-frequency data by summing daily log-returns for each d window size, and this is equivalent to sub-sampling based on the price data. The variance of the resulting d-day return data is $d \times \sigma^2$. Therefore, we can compare the estimator $\hat{\sigma}_h/d$ with the true variance σ^2, where $\hat{\sigma}_h = T^{-\alpha}\sum_{t=1}^{T^\alpha}(y_t - \bar{y})^2$ using d-day log-returns. Using this fact and Assumption 3.1(iii), we can impose the above element-wise convergence condition. However, Structured-POET does not require this assumption because it can remove the scale issue by using the correlation matrix and recovering with daily-based variance estimator $\hat{D}$ in Section 2.2. In the simulation study, we used $d^{-1}\widehat{\boldsymbol{\Sigma}}_h$ for the initial sample covariance matrix.

Similar to the proofs of Choi and Kim (2023), we can show that Double-POET yields the following convergence rates.

Theorem 3.2. *Suppose that $m_p = o(p^{c(5a_2-3)/2})$ and Assumptions 3.1–3.2 hold. Let $\omega_{T^\alpha} = p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5}{2}a_1-\frac{3}{2}+c(\frac{5}{2}a_2-\frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}$. If $m_p\omega_{T^\alpha}^{1-q} = o(1)$, we have*

$$\|\widehat{\mathbf{\Sigma}}^{\mathcal{D}} - \mathbf{\Sigma}\|_{\max} = O_P\left(\omega_{T^\alpha}\right), \tag{3.4}$$

$$\|(\widehat{\mathbf{\Sigma}}^{\mathcal{D}})^{-1} - \mathbf{\Sigma}^{-1}\|_2 = O_P\left(m_p\omega_{T^\alpha}^{1-q} + p^{\frac{c}{2}(1-a_2)}\omega_{T^\alpha} + p^{3(1-a_1)}\Big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{3a_1-2-c}}\right). \tag{3.5}$$

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

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}^{\mathcal{D}} - \mathbf{\Sigma}\|_{\Sigma} =& O_P\Big(m_p\omega_{T^\alpha}^{1-q} + p^{\frac{11}{2}-5a_1+5c(1-a_2)}\big(\frac{\log p}{T^\alpha} + \frac{1}{T^{2(1-\alpha)\beta}}\big)\\ &+ \frac{1}{p^{5a_1-\frac{7}{2}-c(7-5a_2)}} + \frac{m_p^2}{p^{5ca_2-3c-\frac{1}{2}}}\Big).\end{aligned} \tag{3.6}$$

We compare the rates of convergence between Structured-POET and Double-POET as follows. For simplicity, consider $m_p = O(1)$, $a_1 = 1$, and $a_2 = 1$, and ignore the log order terms. Define the optimal $\alpha^* = \frac{2\beta}{1+2\beta}$ (see Remark 3.2). Under the relative Frobenius norm with $\alpha = \alpha^*$, we have

$$\|\widehat{\mathbf{\Sigma}}^{\mathcal{D}} - \mathbf{\Sigma}\|_{\Sigma} = O_P\left(\left(\frac{1}{T^{\frac{\beta}{1+2\beta}}} + \frac{1}{p^{1-c}} + \frac{1}{p^c}\right)^{1-q} + \frac{\sqrt{p}}{T^{\frac{2\beta}{1+2\beta}}} + \frac{1}{p^{\frac{3}{2}-2c}} + \frac{1}{p^{2c-\frac{1}{2}}}\right),$$

$$\|\widehat{\mathbf{\Sigma}}^{\mathcal{S}} - \mathbf{\Sigma}\|_{\Sigma} = O_P\left(\left(\frac{1}{\sqrt{T}} + \frac{1}{p^{1-c}} + \frac{1}{p^c}\right)^{1-q} + \frac{1}{T^{\frac{\beta}{1+2\beta}}} + \frac{\sqrt{p}}{T^{\frac{2\beta}{1+2\beta}}} + \frac{1}{p^{\frac{3}{2}-2c}} + \frac{1}{p^{2c-\frac{1}{2}}}\right).$$

Specifically, when $q \neq 0$, Structured-POET achieves a faster convergence rate under the relative Frobenius norm. This is because utilizing all observations enhances the estimation accuracy of each block diagonal matrix. However, when $q = 0$, the convergence rates of both estimators are the same. This is because the estimation error of the correlations between continents dominates the benefit mentioned above. Importantly, we note that this does not mean that their estimation errors are exactly the same. In fact, based on our simulation study, we can conjecture that Structured-POET has smaller convergence rates than Double-POET for $q = 0$. That is, the relative ratio of the convergence rate of Structured-POET with respect to that of Double-POET may be less than 1. Unfortunately, due to the complex upper bound calculations used to handle high-dimensional matrices, we cannot theoretically show this statement for $q = 0$. We leave this for a future study. Similarly, under the spectral norm for the inverse matrix, the convergence rate of Structure-POET can be faster than that of Double-POET when $q \neq 0$.

4 Simulation Study

In this section, simulations are carried out to examine the finite sample performance of the proposed Structured-POET method. We considered the true covariance as $\Sigma = BB' + \Lambda\Lambda' + \Sigma_u$, where each row of B was drawn from $\mathcal{N}(\mu_B, I_k)$, where each element of μ_B is i.i.d. Uniform$(-0.5, 0.5)$; for $\Lambda = \mathrm{diag}(\Lambda^1, \dots, \Lambda^L)$, each row of Λ^l for each l was drawn from $\mathcal{N}(\mu_{\Lambda^l}, I_{r_l})$, where each element of μ_{Λ^l} is i.i.d. Uniform$(-0.3, 0.3)$. We generated Σ_u as follows. Let $D_u = \mathrm{diag}(d_1, \dots, d_p)$, where each $\{d_i\}$ was generated independently from Uniform$(0.5, 1.5)$. Let $\pi = (\pi_1, \dots, \pi_p)'$ be a sparse vector, where each π_i was drawn from $\mathcal{N}(0, 1)$ with probability $\frac{0.5}{\sqrt{p}\log p}$, and $\pi_i = 0$ otherwise. Then, we set $\Sigma_u = D_u + \pi\pi' - \mathrm{diag}\{\pi_1^2, \dots, \pi_p^2\}$. In the simulation, we generated Σ_u until it was positive definite.

Let D be the diagonal matrix consisting of the diagonal elements of Σ. We then obtained the true correlation matrix $R = D^{-\frac{1}{2}}\Sigma D^{-\frac{1}{2}} = (\rho_{0,ij})_{p\times p}$. Next, we set $R_h = (\rho_{h,ij})_{p\times p}$, where $\rho_{h,ij} = \mathrm{sgn}(\rho_{0,ij})(|\rho_{0,ij}| + 0.5h^{\beta})$ if i and j belong to different continent groups, for $h = \frac{0.5}{d}$ and $\beta = 0.75$, and $\rho_{h,ij} = \rho_{0,ij}$ if i and j are in the same continent group. Let $\{\gamma_i, v_i\}_{i=1}^k$ be the leading eigenvalues and eigenvectors of $\widetilde{\Sigma}_g = D^{\frac{1}{2}} R_h D^{\frac{1}{2}} - \Lambda\Lambda' - \Sigma_u$. Then, we obtained $B_h = V\Gamma^{\frac{1}{2}}$, where $\Gamma = \text{diag}(\gamma_1, \ldots, \gamma_k)$ and $V = (v_1, \ldots, v_k)$. We note that B_h represents the non-synchronized structure. Thus, we generated non-synchronized observations by

$$y_t = B_h G_t + \Lambda F_t + u_t,$$

where G_t, F_t, and u_t were drawn from $\mathcal{N}(0, I_k)$, $\mathcal{N}(0, I_r)$, and $\mathcal{N}(0, \Sigma_u)$, respectively.

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, Double-POET (D-POET), and Structured-POET (S-POET) methods were employed to estimate $\mathbf{\Sigma}$. The average estimation errors were measured under the following norms: $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$, $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max}$, and $\|(\widehat{\mathbf{\Sigma}})^{-1} - \mathbf{\Sigma}^{-1}\|$, where $\widehat{\mathbf{\Sigma}}$ is one of the covariance matrix estimators. We note that the lower-frequency data is 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 $d^{-1}\widehat{\mathbf{\Sigma}}_h$ (see Section 3.1). 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 the S-POET estimation, 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

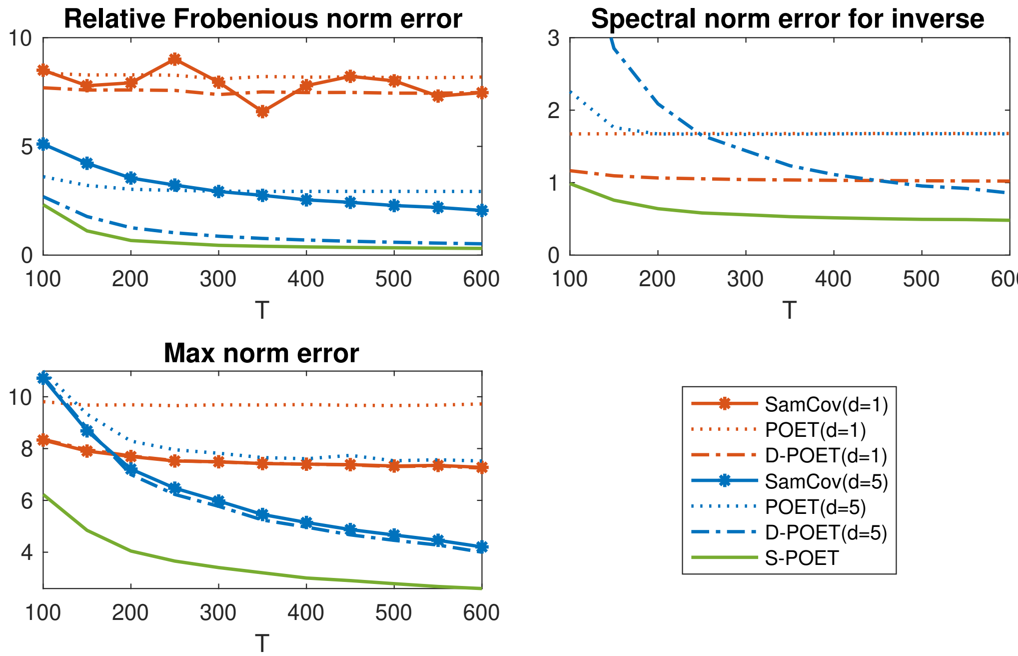

Figure 1: Averages of $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$, $\|(\widehat{\mathbf{\Sigma}})^{-1} - \mathbf{\Sigma}^{-1}\|$, and $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max}$ for the sample covariance matrix, POET, Double-POET, and Structured-POET against T with fixed $p = 500$ and $L = 20$. Lines that exceed the upper limits of the y-axis are excluded.

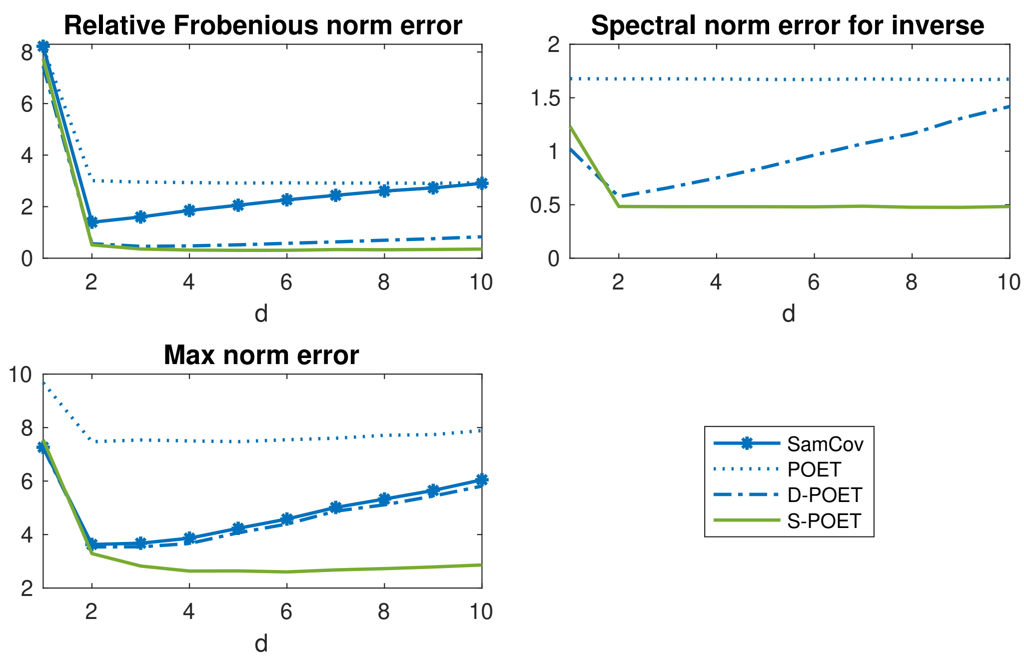

Figure 2: Averages of $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$, $\|(\widehat{\mathbf{\Sigma}})^{-1} - \mathbf{\Sigma}^{-1}\|$, and $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max}$ for the sample covariance matrix, POET, Double-POET, and Structured-POET against d with fixed $p = 500$, $T = 600$, and $L = 20$. Lines that exceed the upper limits of the y-axis are excluded.

T and d, respectively. From Figures 1 and 2, we find that S-POET has smaller estimation errors than the other methods under different norms. 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 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 $d = 4$. In contrast, as the frequency size d increases, the estimation errors of SamCov 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, the proposed 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 the proposed Structured-POET method with global financial data. We obtained the daily transaction prices of international stock markets over 15 countries by the total market capitalization. The whole

Table 1: Distributions of the number of firms

America		Asia		Europe	
United States (US)	221	China (CN)	100	United Kingdom (GB)	100
Canada (CA)	100	Japan (JP)	100	France (FR)	100
Brazil (BR)	100	Hong Kong (HK)	100	Germany (DE)	100
Mexico (MX)	48	India (IN)	100	Switzerland (CH)	100
Chile (CL)	31	Korea (KR)	100	Sweden (SE)	100
				Total	1500

sample period is from January 3, 2017, to December 30, 2022. After excluding stocks with missing returns and no variation, we picked 1500 stocks for this period 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 our sample is presented in Table 1.

We computed several estimators, including Structured-POET (S-POET), Double-POET (D-POET), POET, and the sample covariance matrix (SamCov) estimators, for each week. In the case of the S-POET procedure, we used weekly returns to estimate the global factor component and daily returns to estimate the local factor and idiosyncratic components. We employed both daily 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 (Ait-Sahalia and Xiu, 2017; Fan et al., 2016). 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 both D-POET and POET. For D-POET, we chose the number of local factors using the eigenvalue ratio method proposed by Ahn and Horenstein (2013) with $r_{l,\max} = 5$. In the S-POET procedure, for each off-diagonal partitioned block, we employed the best rank-one approximation as suggested by the largest singular value gap method mentioned in Remark 2.1.

We considered the following constrained minimum variance portfolio allocation problem (Fan et al., 2012; Jagannathan and Ma, 2003) to analyze the out-of-sample portfolio

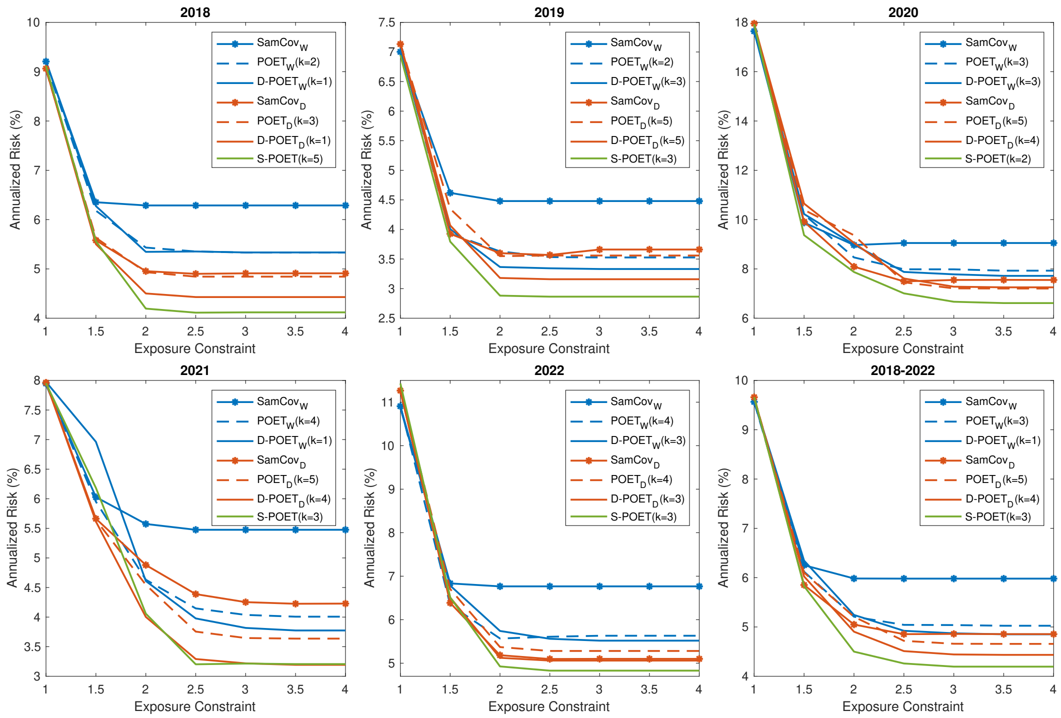

Figure 3: Out-of-sample risks of the optimal portfolios constructed by the SamCov, POET, Double-POET, and Structured-POET estimators for the global stock market.

allocation performance:

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

where $\mathbf{1} = (1, \ldots, 1)^\top \in \mathbb{R}^p$, the gross exposure constraint c varies from 1 to 4, and $\widehat{\boldsymbol{\Sigma}}$ 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 one 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 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 $k = 1, \ldots, 5$ for each estimator type and each period. We used subscripts to explicitly denote the frequency of the data used, with W (blue lines) and D (red lines) representing weekly and daily data, respectively. As shown in Figure 3, S-POET consistently outperforms the other estimators. Specifically, for all periods except 2021, S-POET 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 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 Double-POET, which use only daily 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 paper, we introduce a novel large global volatility matrix inference procedure. The proposed Structured-POET method leverages observation structural information from global financial markets based on latent global and national factor models. We establish the asymptotic properties of Structured-POET and demonstrate its efficiency in estimating a large global covariance matrix compared to the Double-POET procedure.

In our empirical study, we demonstrate that the proposed Structured-POET estimator outperforms the other existing methods in minimum variance portfolio allocation problems. This is because the Structured-POET procedure 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. Overall, our findings support the effectiveness of the Structured-POET method in estimating large global volatility matrices.

References

Ahn, S. C. and A. R. Horenstein (2013): "Eigenvalue ratio test for the number of factors," *Econometrica*, 81, 1203–1227.

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

Ando, T. and J. Bai (2015): "Asset pricing with a general multifactor structure," *Journal of Financial Econometrics*, 13, 556–604.

——— (2016): "Panel data models with grouped factor structure under unknown group membership," *Journal of Applied Econometrics*, 31, 163–191.

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

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

Bekaert, G., R. J. Hodrick, and X. Zhang (2009): "International stock return co-movements," *The Journal of Finance*, 64, 2591–2626.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly journal of economics*, 120, 387–422.

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

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

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

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

Choi, I., D. Kim, Y. J. Kim, and N.-S. Kwark (2018): "A multilevel factor model: Identification, asymptotic theory and applications," *Journal of Applied Econometrics*, 33, 355–377.

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

Fama, E. F. and K. R. French (2012): "Size, value, and momentum in international stock returns," *Journal of financial economics*, 105, 457–472.

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

Fan, J., Q. Li, and Y. Wang (2017): "Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 247–265.

Fan, J., Y. Liao, and M. Mincheva (2011): "High dimensional covariance matrix estimation in approximate factor models," *Annals of statistics*, 39, 3320.

——— (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75.

Fan, J., H. Liu, and W. Wang (2018a): "Large covariance estimation through elliptical factor models," *Annals of statistics*, 46, 1383.

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

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

Fan, J., J. Zhang, and K. Yu (2012): "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.

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

Jagannathan, R. and T. Ma (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps," *The Journal of Finance*, 58, 1651–1683.

Jung, K., D. Kim, and S. Yu (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., C. Otrok, and C. H. Whiteman (2003): "International business cycles: World, region, and country-specific factors," *american economic review*, 93, 1216–1239.

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

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

Onatski, A. (2010): "Determining the number of factors from empirical distribution of eigenvalues," *The Review of Economics and Statistics*, 92, 1004–1016.

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

Stein, C. and W. James (1961): "Estimation with quadratic loss," in *Proc. 4th Berkeley Symp. Mathematical Statistics Probability*, vol. 1, 361–379.

Stock, J. H. and M. W. Watson (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American statistical association*, 97, 1167–1179.

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

WANG, W. AND J. FAN (2017): "Asymptotics of empirical eigenstructure for high dimensional spiked covariance," *Annals of statistics*, 45, 1342.

A Appendix

A.1 Double-POET procedure

We decompose the covariance matrix of the sth continent as follows:

$$\boldsymbol{\Sigma}^s = \boldsymbol{\Sigma}_g^s + \boldsymbol{\Sigma}_l^s + \boldsymbol{\Sigma}_u^s.$$

Then, each component as well as $\boldsymbol{\Sigma}^s$ can be estimated by the Double-POET procedure (Choi and Kim, 2023) as follows:

1. Given a sample covariance matrix, $\widehat{\boldsymbol{\Sigma}}^s$, using T observations, let $\{\widehat{\delta}_i^s, \widehat{v}_i^s\}_{i=1}^p$ be the eigenvalues and eigenvectors of $\widehat{\boldsymbol{\Sigma}}^s$ in decreasing order. We compute

   $$\widehat{\boldsymbol{\Sigma}}_g^{s,\mathcal{D}} = \widehat{\mathbf{V}}^s \widehat{\boldsymbol{\Gamma}}^s \widehat{\mathbf{V}}^{s\prime},$$

   where $\widehat{\boldsymbol{\Gamma}}^s = \mathrm{diag}(\widehat{\delta}_1^s, \ldots, \widehat{\delta}_k^s)$ and $\widehat{\mathbf{V}}^s = (\widehat{v}_1^s, \ldots, \widehat{v}_k^s)$.

2. Define $\widehat{\boldsymbol{\Sigma}}_E^{l,s}$ as each $p_l \times p_l$ diagonal block of $\widehat{\boldsymbol{\Sigma}}_E^s = \widehat{\boldsymbol{\Sigma}}^s - \widehat{\boldsymbol{\Sigma}}_g^{s,\mathcal{D}}$. For the lth block, let $\{\widehat{\kappa}_i^{l,s}, \widehat{\eta}_i^{l,s}\}_{i=1}^{p_l}$ be the eigenvalues and eigenvectors of $\widehat{\boldsymbol{\Sigma}}_E^{l,s}$ in decreasing order. Then, we compute

   $$\widehat{\boldsymbol{\Sigma}}_l^{s,\mathcal{D}} = \widehat{\boldsymbol{\Phi}}^s \widehat{\boldsymbol{\Psi}}^s \widehat{\boldsymbol{\Phi}}^{s\prime},$$

   where $\widehat{\boldsymbol{\Psi}}^s = \mathrm{diag}(\widehat{\Psi}^1, \ldots, \widehat{\Psi}^{L_s})$ for $\widehat{\Psi}^l = \mathrm{diag}(\widehat{\kappa}_1^{l,s}, \ldots, \widehat{\kappa}_{r_l}^{l,s})$, and the block diagonal matrix $\widehat{\boldsymbol{\Phi}}^s = \mathrm{diag}(\widehat{\Phi}^1, \ldots, \widehat{\Phi}^{L_s})$ for $\widehat{\Phi}^l = (\widehat{\eta}_1^{l,s}, \ldots, \widehat{\eta}_{r_l}^{l,s})$ for $l = 1, 2, \ldots, L_s$, where L_s is the number of countries in continent s.

3. Let $\widehat{\boldsymbol{\Sigma}}_u^s = \widehat{\boldsymbol{\Sigma}}^s - \widehat{\boldsymbol{\Sigma}}_g^{s,\mathcal{D}} - \widehat{\boldsymbol{\Sigma}}_l^{s,\mathcal{D}}$ be the principal orthogonal complement. We apply the adaptive thresholding method on $\widehat{\boldsymbol{\Sigma}}_u^s = (\widehat{\sigma}_{u,ij})_{p\times p}$ following Bickel and Levina (2008) and Fan et al. (2013). Specifically, define $\widehat{\boldsymbol{\Sigma}}_u^{s,\mathcal{D}}$ as the thresholded error covariance

matrix estimator:

$$\widehat{\boldsymbol{\Sigma}}_u^{s,\mathcal{D}} = (\widehat{\sigma}_{u,ij}^{s,\mathcal{D}})_{p_s \times p_s}, \qquad \widehat{\sigma}_{u,ij}^{s,\mathcal{D}} = \begin{cases} \widehat{\sigma}_{u,ij}, & i = j \\ s_{ij}(\widehat{\sigma}_{u,ij}) I(|\widehat{\sigma}_{u,ij}| \geq \tau_{ij}), & i \neq j \end{cases},$$

where an entry-dependent threshold $\tau_{ij} = \tau(\widehat{\sigma}_{u,ii}\widehat{\sigma}_{u,jj})^{1/2}$ and $s_{ij}(\cdot)$ is a generalized thresholding function (e.g., hard or soft thresholding; see Cai and Liu, 2011; Rothman et al., 2009). The thresholding constant is determined by $\tau \asymp \omega_T$, where ω_T is defined in Theorem 3.1.

4. The final estimator of $\boldsymbol{\Sigma}^s$ is then defined as

$$\widehat{\boldsymbol{\Sigma}}^{s,\mathcal{D}} = \widehat{\boldsymbol{\Sigma}}_g^{s,\mathcal{D}} + \widehat{\boldsymbol{\Sigma}}_l^{s,\mathcal{D}} + \widehat{\boldsymbol{\Sigma}}_u^{s,\mathcal{D}}.$$

By using the proof of Theorem 3.1 of Choi and Kim (2023) and Assumption 3.1, we can obtain the following results: for each continent $s \in \{1, \ldots, S\}$,

$$\|\widehat{\boldsymbol{\Sigma}}_g^{s,\mathcal{D}} - \boldsymbol{\Sigma}_g^s\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)}\sqrt{\log p/T} + 1/p^{\frac{5a_1}{2} - \frac{3}{2} - c}), \tag{A.1}$$

$$\|\widehat{\boldsymbol{\Sigma}}_l^{s,\mathcal{D}} - \boldsymbol{\Sigma}_l^s\|_{\max} = O_P(\omega_T), \tag{A.2}$$

$$\|\widehat{\boldsymbol{\Sigma}}_u^{s,\mathcal{D}} - \boldsymbol{\Sigma}_u^s\|_{\max} = O_P(\omega_T), \tag{A.3}$$

where $\omega_T = p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}$.

A.2 Proof of Theorem 3.1

We first provide useful lemmas below. Let $\{\delta_i, v_i\}_{i=1}^p$ be the eigenvalues and their corresponding eigenvectors of $\boldsymbol{\Sigma}$ in decreasing order. 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 $\mathbf{BB}'$ and $\widetilde{\boldsymbol{\Sigma}}_g$, respectively, where $\widetilde{\boldsymbol{\Sigma}}_g = (\widetilde{\boldsymbol{\Sigma}}_g^{\mathcal{D}} + \hat{D}^{\frac{1}{2}}\widehat{\boldsymbol{\Theta}}\hat{D}^{\frac{1}{2}})$. Define $\boldsymbol{\Sigma}_E = \boldsymbol{\Lambda\Lambda}' + \boldsymbol{\Sigma}_u$ and let $\boldsymbol{\Sigma}_E^l = \Lambda^l\Lambda^{l\prime} + \boldsymbol{\Sigma}_u^l$ be the lth diagonal block of $\boldsymbol{\Sigma}_E$. For each country l, let $\{\kappa_i^l, \eta_i^l\}_{i=1}^{p_l}$ be the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}_E^l$ in decreasing order, and

$\{\bar{\kappa}_i^l, \bar{\eta}_i^l\}_{i=1}^{r_l}$ for $\Lambda^l \Lambda^{l\prime}$.

By Weyl's theorem, we have the following lemma under the pervasive conditions.

Lemma A.1. *Under Assumption 3.1(i), we have*

$$|\delta_i - \bar{\delta}_i| \leq \|\mathbf{\Sigma}_E\| \text{ for } i \leq k, \qquad |\delta_i| \leq \|\mathbf{\Sigma}_E\| \text{ for } i > k,$$

and, for $i \leq k$, $\bar{\delta}_i/p^{a_1}$ is strictly bigger than zero for all p. In addition, for each national group l, we have

$$|\kappa_i^l - \bar{\kappa}_i^l| \leq \|\mathbf{\Sigma}_u^l\| \text{ for } i \leq r_l, \qquad |\kappa_i^l| \leq \|\mathbf{\Sigma}_u^l\| \text{ for } i > r_l,$$

and, for $i \leq r_l$, $\bar{\kappa}_i^l/p_l^{a_2}$ is strictly bigger than zero for all p_l.

The following lemma presents the individual convergence rate of leading eigenvectors using Lemma A.1 and the l_∞ norm perturbation bound theorem of Fan et al. (2018b).

Lemma A.2. *Under Assumption 3.1(i), we have the following results.*

(i) We have

$$\max_{i \leq k} \|\bar{v}_i - v_i\|_\infty \leq C \frac{\|\mathbf{\Sigma}_E\|_\infty}{p^{3(a_1 - \frac{1}{2})}}.$$

(ii) For each national group l, we have

$$\max_{i \leq r_l} \|\bar{\eta}_i^l - \eta_i^l\|_\infty \leq C \frac{\|\mathbf{\Sigma}_u^l\|_\infty}{p_l^{3(a_2 - \frac{1}{2})}}.$$

Proof. (i) Let $\mathbf{B} = (\tilde{b}_1, \ldots, \tilde{b}_k)$. Then, for $i \leq k$, $\bar{\delta}_i = \|\tilde{b}_i\|^2 \asymp p^{a_1}$ from Lemma A.1 and $\bar{v}_i = \tilde{b}_i/\|\tilde{b}_i\|$. Hence, $\|\bar{v}_i\|_\infty \leq \|\mathbf{B}\|_{\max}/\|\tilde{b}_i\| \leq C/\sqrt{p^{a_1}}$. In addition, for $\widetilde{\mathbf{V}} = (\bar{v}_1, \ldots, \bar{v}_k)$, the coherence $\mu(\widetilde{\mathbf{V}}) = p \max_i \sum_{j=1}^k \widetilde{\mathbf{V}}_{ij}^2/k \leq Cp^{1-a_1}$, where $\widetilde{\mathbf{V}}_{ij}$ is the (i,j) entry of $\widetilde{\mathbf{V}}$. Thus, by

Theorem 1 of Fan et al. (2018b), we have

$$\max_{i\le k}\|\bar{v}_i - v_i\|_\infty \le Cp^{2(1-a_1)}\frac{\|\mathbf{\Sigma}_E\|_\infty}{\bar{\gamma}\sqrt{p}},$$

where the eigengap $\bar{\gamma} = \min\{\bar{\delta}_i - \bar{\delta}_{i+1} : 1 \le i \le k\}$ and $\delta_{k+1} = 0$. By the similar argument, we can show the result (ii). $\square$

Lemma A.3. *Let $\mathbf{R}_0 = (R_{0,sq})_{S\times S}$, where $R_{0,sq}$ is the (s,q)th off-diagonal partitioned block matrix for $s, q \in \{1,\dots,S\}$. Under Assumption 3.1, for $s \ne q$, we have*

$$\|\widehat{\Theta}_{sq} - R_{0,sq}\|_{\max} = O_P\left(p^{\frac{5}{2}(1-a_1)}\big(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}\big)\right).$$

Proof. For $s \ne q$, let the singular value decomposition be $R_{0,sq} = \mathbf{U}\Xi\mathbf{W}' = \sum_{i=1}^{k^*_{sq}} \xi_i u_i w_i'$, where k^*_{sq} is the rank of $R_{0,sq}$, the singular values are $\xi_1 \ge \xi_2 \ge \cdots \ge \xi_{k^*_{sq}} > 0$, and the matrices $\mathbf{U} = (u_1,\dots,u_{k^*_{sq}})$, $\mathbf{W} = (w_1,\dots,w_{k^*_{sq}})$ consist of the singular vectors. By Lipschitz condition, $\|\mathbf{R}_h - \mathbf{R}_0\|_{\max} = O(1/T^{(1-\alpha)\beta})$, and Assumption 3.1 (iii), we have

$$\|\widehat{\mathbf{R}}_h - \mathbf{R}_0\|_{\max} = O_P\left(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\right). \tag{A.4}$$

Note that $R_{0,sq}$ is k^*_{sq}-rank matrix for $s \ne q \in \{1,\dots,S\}$. By Weyl's inequality, we have

$$|\widehat{\xi}_i - \xi_i| \le \|\widehat{R}_{sq} - R_{0,sq}\|_F = O_P\left(p\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big)\right). \tag{A.5}$$

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

$$\begin{aligned}\|\widehat{u}_i - u_i\|_\infty &\le Cp^{2(1-a_1)}\frac{\|\widehat{R}_{sq} - R_{0,sq}\|_\infty}{p^{a_1}\sqrt{p}} \le Cp^{2(1-a_1)}\frac{\|\widehat{R}_{sq} - R_{0,sq}\|_{\max}}{p^{a_1-1}\sqrt{p}} \\ &= O_P\left(p^{\frac{5}{2}-3a_1}\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big)\right).\end{aligned} \tag{A.6}$$

Similarly, we can obtain the same rate for $\|\widehat{w}_i - w_i\|_\infty$. Note that $\|\mathbf{U}\Xi^{\frac{1}{2}}\|_{\max} = O_P(1)$. By

(A.5) and (A.6), we have

$$
\begin{aligned}
\|\widehat{\mathbf{U}}\widehat{\Xi}^{\frac{1}{2}} - \mathbf{U}\Xi^{\frac{1}{2}}\|_{\max} &\leq \|\widehat{\mathbf{U}}(\widehat{\Xi}^{\frac{1}{2}} - \Xi^{\frac{1}{2}})\|_{\max} + \|(\widehat{\mathbf{U}} - \mathbf{U})\Xi^{\frac{1}{2}}\|_{\max} \\
&= O_P\left(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta})\right) = o\,(1)\,.
\end{aligned}
$$

Then, we have $\|\widehat{\mathbf{U}}\|_{\max} = O_P(1/\sqrt{p^{a_1}})$. Similarly, we can obtain $\|\widehat{\mathbf{W}}\|_{\max} = O_P(1/\sqrt{p^{a_1}})$. Therefore, we have

$$
\begin{aligned}
\|\widehat{\Theta}_{sq} - R_{0,sq}\|_{\max} &\leq \|\widehat{\mathbf{U}}(\widehat{\Xi} - \Xi)\widehat{\mathbf{W}}'\|_{\max} + \|(\widehat{\mathbf{U}} - \mathbf{U})\Xi(\widehat{\mathbf{W}} - \mathbf{W})'\|_{\max} + 2\|(\widehat{\mathbf{U}} - \mathbf{U})\Xi\mathbf{W}'\|_{\max} \\
&= O_P(p^{-a_1}\|\widehat{\Xi} - \Xi\|_{\max} + \sqrt{p^{a_1}}\|\widehat{\mathbf{U}} - \mathbf{U}\|_{\max}) = O_P\big(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta})\big).
\end{aligned}
$$

$\square$

Lemma A.4. *Under Assumption 3.1, for $i \leq k$, we have*

$$
\begin{aligned}
|\widetilde{\delta}_i - \bar{\delta}_i| &= O_P\left(p^{\frac{7}{2}-\frac{5}{2}a_1}\Big(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{5}{2}a_1-\frac{5}{2}-c}}\right), \\
\|\widetilde{v}_i - \bar{v}_i\|_{\infty} &= O_P\left(p^{5-\frac{11}{2}a_1}\Big(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{11}{2}a_1-4-c}}\right).
\end{aligned}
$$

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 A.3, we can easily obtain that

$$
\|\hat{D}^{\frac{1}{2}}\widehat{\mathbf{\Theta}}\hat{D}^{\frac{1}{2}} - D^{\frac{1}{2}}\widetilde{\mathbf{R}}_0 D^{\frac{1}{2}}\|_{\max} = O_P\big(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta})\big). \tag{A.7}
$$

By using the fact that $\widetilde{\mathbf{\Sigma}}_g$ and $\mathbf{\Sigma}_g$ are low-rank matrices, (A.1) and (A.7), we have

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

$$= O_P\left(p\Big(p^{\frac{5}{2}(1-a_1)}(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}) + \frac{1}{p^{\frac{5}{2}a_1-\frac{3}{2}-c}}\Big)\right).$$

By Theorem 1 of Fan et al. (2018b), (A.1) and (A.7), we have

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

$\square$

Proof of Theorem 3.1. Consider (3.1). Let $\mathbf{B}\mathbf{B}' = \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}\widetilde{\mathbf{V}}'$, where $\widetilde{\boldsymbol{\Gamma}} = \mathrm{diag}(\bar{\delta}_1, \dots, \bar{\delta}_k)$ and their corresponding leading k eigenvectors $\widetilde{\mathbf{V}} = (\bar{v}_1, \dots, \bar{v}_k)$. By Lemma A.4, we have

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

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

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

Consider (3.2). Similar to the proofs of (A.22), we can show $\|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1} - \boldsymbol{\Sigma}_E^{-1}\| = O_P(m_p\omega_T^{1-q} + p^{\frac{c}{2}(1-a_2)}\omega_T)$. Let $\widehat{\mathbf{H}} = \widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}}\widetilde{\mathbf{V}}_g'(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1}\widetilde{\mathbf{V}}_g\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}}$ and $\widetilde{\mathbf{H}} = \widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'\boldsymbol{\Sigma}_E^{-1}\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}$. Using the Sherman-Morrison-Woodbury formula, we have

$$\|(\widehat{\boldsymbol{\Sigma}}^{\mathcal{S}})^{-1} - \boldsymbol{\Sigma}^{-1}\| \le \|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1} - \boldsymbol{\Sigma}_E^{-1}\| + \Delta,$$

where $\Delta = \|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1}\widetilde{\mathbf{V}}_g\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}}(\mathbf{I}_k + \widehat{\mathbf{H}})^{-1}\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}}\widetilde{\mathbf{V}}_g'(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1} - \boldsymbol{\Sigma}_E^{-1}\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}(\mathbf{I}_k + \widetilde{\mathbf{H}})^{-1}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'\boldsymbol{\Sigma}_E^{-1}\|$. Then, the right hand side can be bounded by following terms:

$$
\begin{aligned}
L_1 &= \|((\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1} - \boldsymbol{\Sigma}_E^{-1})\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}(\mathbf{I}_k + \widetilde{\mathbf{H}})^{-1}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'\boldsymbol{\Sigma}_E^{-1}\|, \\
L_2 &= \|\boldsymbol{\Sigma}_E^{-1}(\widetilde{\mathbf{V}}_g\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}} - \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}})(\mathbf{I}_k + \widetilde{\mathbf{H}})^{-1}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'\boldsymbol{\Sigma}_E^{-1}\|, \\
L_3 &= \|\boldsymbol{\Sigma}_E^{-1}\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}((\mathbf{I}_k + \widehat{\mathbf{H}})^{-1} - (\mathbf{I}_k + \widetilde{\mathbf{H}})^{-1})\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'\boldsymbol{\Sigma}_E^{-1}\|.
\end{aligned}
$$

By Weyl's inequality, we have $\lambda_{\min}(\boldsymbol{\Sigma}_E) > c$ since $\lambda_{\min}(\boldsymbol{\Sigma}_u) > c$ and $\lambda_{\min}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}') = 0$. Hence, $\|\boldsymbol{\Sigma}_E^{-1}\| = O_P(1)$. Note that $\|\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\| = O_P(p^{\frac{a_1}{2}})$. By Lemma A.4, we have $\|\widetilde{\mathbf{V}}_g\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}} - \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\|_{\max} = O_P(p^{5(1-a_1)}\big(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}\big) + 1/p^{5a_1-4-c})$. Then, we have

$$
\begin{aligned}
\|\widehat{\mathbf{H}} - \widetilde{\mathbf{H}}\| &\leq \|(\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}}\widetilde{\mathbf{V}}_g' - \widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}')(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1}(\widetilde{\mathbf{V}}_g\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}} - \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}})\| \\
&\quad + \|(\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}}\widetilde{\mathbf{V}}_g' - \widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}')(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1}\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\| + \|\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'((\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{S}})^{-1} - \boldsymbol{\Sigma}_E^{-1})\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\| \\
&= O_P\left(p^{a_1}m_p\omega_T^{1-q} + p^{\frac{11}{2}-\frac{9}{2}a_1}\left(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\right) + \frac{1}{p^{\frac{9}{2}(a_1-1)-c}}\right).
\end{aligned}
$$

Since $\lambda_{\min}(\mathbf{I}_k + \widetilde{\mathbf{H}}) \geq \lambda_{\min}(\widetilde{\mathbf{H}}) \geq \lambda_{\min}(\boldsymbol{\Sigma}_E^{-1})\lambda_{\min}^2(\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}) \geq Cp^{a_1}$, we have $\|(\mathbf{I}_k + \widetilde{\mathbf{H}})^{-1}\| = O_P(1/p^{a_1})$. Then, $L_1 = O_P(m_p\omega_T^{1-q})$. In addition, $L_2 = O_P(p^{\frac{-a_1}{2}}\|\widetilde{\mathbf{V}}_g\widetilde{\boldsymbol{\Gamma}}_g^{\frac{1}{2}} - \widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\|) = O_P(p^{\frac{11}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{11}{2}a_1-\frac{9}{2}-c})$ and $L_3 = O_P(p^{a_1}\|(\mathbf{I}_k + \widehat{\mathbf{H}})^{-1} - (\mathbf{I}_k + \widetilde{\mathbf{H}})^{-1}\|) = O_P(p^{-a_1}\|\widehat{\mathbf{H}} - \widetilde{\mathbf{H}}\|) = O_P(m_p\omega_T^{1-q} + p^{\frac{11}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{11}{2}a_1-\frac{9}{2}-c})$. Thus, we have

$$
\Delta = O_P(m_p\omega_T^{1-q} + p^{\frac{11}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{11}{2}a_1-\frac{9}{2}-c}). \tag{A.8}
$$

Therefore, we have

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

Consider (3.3). We derive the rate of convergence for $\|\widehat{\boldsymbol{\Sigma}}^{\mathcal{S}} - \boldsymbol{\Sigma}\|_{\Sigma}$. The SVD decomposition of $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = \begin{pmatrix}\mathbf{V}_{p\times k} & \mathbf{\Phi}_{p\times r} & \mathbf{\Omega}_{p\times(p-k-r)}\end{pmatrix}\begin{pmatrix}\mathbf{\Gamma}_{k\times k} & & \\ & \mathbf{\Psi}_{r\times r} & \\ & & \mathbf{\Theta}_{(p-k-r)\times(p-k-r)}\end{pmatrix}\begin{pmatrix}\mathbf{V}' \\ \mathbf{\Phi}' \\ \mathbf{\Omega}'\end{pmatrix}.$$

Note that $\mathbf{\Omega}$ is used to denote the precision matrix in Section 2.2. Moreover, since all the eigenvalues of $\mathbf{\Sigma}$ are strictly bigger than 0, for any maxtrix $\mathbf{A}$, we have $\|\mathbf{A}\|_\Sigma^2 = O_P(p^{-1})\|\mathbf{A}\|_F^2$. Then, we have

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}^{\mathcal{S}} - \mathbf{\Sigma}\|_\Sigma &\le p^{-1/2}\Big(\|\mathbf{\Sigma}^{-1/2}(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\mathbf{\Sigma}^{-1/2}\|_F \\ &\qquad + \|\mathbf{\Sigma}^{-1/2}(\widetilde{\mathbf{\Sigma}}_l^{\mathcal{D}} - \mathbf{\Lambda}\mathbf{\Lambda}')\mathbf{\Sigma}^{-1/2}\|_F + \|\mathbf{\Sigma}^{-1/2}(\widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u)\mathbf{\Sigma}^{-1/2}\|_F\Big) \\ &=: \Delta_G + \Delta_L + \Delta_S.\end{aligned}$$

By using the fact that S is fixed and proofs of (3.5) in Choi and Kim (2023), we can obtain

$$\begin{aligned}\Delta_L = O_P\Big(&p^{\frac{5}{2}(1-a_1)+c(1-a_2)}\sqrt{\frac{\log p}{T}} + \frac{1}{p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2}} + \frac{m_p}{p^{ca_2}} \\ &+ p^{\frac{11}{2}-5a_1+5c(1-a_2)}\frac{\log p}{T} + \frac{1}{p^{5a_1-\frac{7}{2}-c(7-5a_2)}} + \frac{m_p^2}{p^{5ca_2-3c-\frac{1}{2}}}\Big)\end{aligned} \tag{A.10}$$

and

$$\Delta_S = O_P(p^{-1/2}\|\widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u\|_F) = O_P(\|\widetilde{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u\|_2) = O_P(m_p\omega_T^{1-q}). \tag{A.11}$$

We have

$$\Delta_G = p^{-1/2}\left\|\begin{pmatrix}\mathbf{\Gamma}^{-\frac{1}{2}}\mathbf{V}' \\ \mathbf{\Psi}^{-\frac{1}{2}}\mathbf{\Phi}' \\ \mathbf{\Theta}^{-\frac{1}{2}}\mathbf{\Omega}'\end{pmatrix}(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\begin{pmatrix}\mathbf{V}\mathbf{\Gamma}^{-\frac{1}{2}} & \mathbf{\Phi}\mathbf{\Psi}^{-\frac{1}{2}} & \mathbf{\Omega}\mathbf{\Theta}^{-\frac{1}{2}}\end{pmatrix}\right\|_F$$

$$
\begin{aligned}
&\leq p^{-1/2}\big(\|\mathbf{\Gamma}^{-1/2}\mathbf{V}'(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\mathbf{V}\mathbf{\Gamma}^{-1/2}\|_F + \|\mathbf{\Psi}^{-1/2}\mathbf{\Phi}'(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\mathbf{\Phi}\mathbf{\Psi}^{-1/2}\|_F \\
&\quad + \|\mathbf{\Theta}^{-1/2}\mathbf{\Omega}'(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\mathbf{\Omega}\mathbf{\Theta}^{-1/2}\|_F + 2\|\mathbf{\Gamma}^{-1/2}\mathbf{V}'(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\mathbf{\Phi}\mathbf{\Psi}^{-1/2}\|_F \\
&\quad + 2\|\mathbf{\Gamma}^{-1/2}\mathbf{V}'(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\mathbf{\Omega}\mathbf{\Theta}^{-1/2}\|_F + 2\|\mathbf{\Psi}^{-1/2}\mathbf{\Phi}'(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{B}\mathbf{B}')\mathbf{\Omega}\mathbf{\Theta}^{-1/2}\|_F\big) \\
&=: \Delta_{G1} + \Delta_{G2} + \Delta_{G3} + 2\Delta_{G4} + 2\Delta_{G5} + 2\Delta_{G6}.
\end{aligned}
$$

In order to find the convergence rate of relative Frobenius norm, we consider the above terms separately. Note that $\mathbf{\Gamma} = \mathrm{diag}(\delta_1, \ldots, \delta_k)$ and $\mathbf{V} = (v_1, \ldots, v_k)$. For Δ_{G1}, we have

$$
\begin{aligned}
\Delta_{G1} &\leq p^{-1/2}\left(\|\mathbf{\Gamma}^{-1/2}\mathbf{V}'(\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g' - \mathbf{V}\mathbf{\Gamma}\mathbf{V}')\mathbf{V}\mathbf{\Gamma}^{-1/2}\|_F + \|\mathbf{\Gamma}^{-1/2}\mathbf{V}'(\mathbf{V}\mathbf{\Gamma}\mathbf{V}' - \mathbf{B}\mathbf{B}')\mathbf{V}\mathbf{\Gamma}^{-1/2}\|_F\right) \\
&=: \Delta_{G1}^{(a)} + \Delta_{G1}^{(b)}.
\end{aligned}
$$

We bound the two terms separately. We have

$$
\begin{aligned}
\Delta_{G1}^{(a)} &\leq p^{-1/2}\big(\|\mathbf{\Gamma}^{-1/2}(\mathbf{V}'\widetilde{\mathbf{V}}_g - \mathbf{I})\widetilde{\mathbf{\Gamma}}_g(\widetilde{\mathbf{V}}_g'\mathbf{V} - \mathbf{I})\mathbf{\Gamma}^{-1/2}\|_F + 2\|\mathbf{\Gamma}^{-1/2}(\mathbf{V}'\widetilde{\mathbf{V}}_g - \mathbf{I})\widetilde{\mathbf{\Gamma}}_g\mathbf{\Gamma}^{-1/2}\|_F \\
&\qquad + \|(\mathbf{\Gamma}^{-1/2}(\widetilde{\mathbf{\Gamma}}_g - \mathbf{\Gamma})\mathbf{\Gamma}^{-1/2}\|_F\big) =: I + II + III.
\end{aligned}
$$

By Lemmas A.2 and A.4, we obtain $\|\mathbf{V}'\widetilde{\mathbf{V}}_g - \mathbf{I}\|_F = \|\mathbf{V}'(\widetilde{\mathbf{V}}_g - \mathbf{V})\|_F \leq \|\widetilde{\mathbf{V}}_g - \mathbf{V}\|_F = O_P(p^{\frac{11}{2}(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{11}{2}a_1 - \frac{9}{2} - c})$. Then, $II = O_P(p^{5-\frac{11}{2}a_1}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{11}{2}a_1 - 4 - c})$ and I is of smaller order. In addition, we have $III \leq \|\mathbf{\Gamma}^{-1/2}(\widetilde{\mathbf{\Gamma}}_g - \mathbf{\Gamma})\mathbf{\Gamma}^{-1/2}\| = O_P(p^{\frac{7}{2}(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{7}{2}a_1 - \frac{5}{2} - c} + 1/p^{a_1 - ca_2})$ by Lemma A.4. Thus, $\Delta_{G1}^{(a)} = O_P(p^{\frac{7}{2}(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{7}{2}a_1 - \frac{5}{2} - c} + 1/p^{a_1 - ca_2})$. Similarly, we have

$$
\begin{aligned}
\Delta_{G1}^{(b)} &\leq p^{-1/2}\big(\|\mathbf{\Gamma}^{-1/2}(\mathbf{V}'\widetilde{\mathbf{V}} - \mathbf{I})\widetilde{\mathbf{\Gamma}}(\widetilde{\mathbf{V}}'\mathbf{V} - \mathbf{I})\mathbf{\Gamma}^{-1/2}\|_F + 2\|\mathbf{\Gamma}^{-1/2}(\mathbf{V}'\widetilde{\mathbf{V}} - \mathbf{I})\widetilde{\mathbf{\Gamma}}\mathbf{\Gamma}^{-1/2}\|_F \\
&\qquad + \|(\mathbf{\Gamma}^{-1/2}(\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma})\mathbf{\Gamma}^{-1/2}\|_F\big) =: I' + II' + III'.
\end{aligned}
$$

By $\sin\theta$ theorem, $\|\mathbf{V}'\widetilde{\mathbf{V}} - \mathbf{I}\| = \|\mathbf{V}'(\widetilde{\mathbf{V}} - \mathbf{V})\| \leq \|\widetilde{\mathbf{V}} - \mathbf{V}\| = O(\|\mathbf{\Sigma}_E\|/p^{a_1})$. Then, we have

$II' = O(1/p^{a_1 - ca_2})$ and I' is of smaller order. By Lemma A.1, we have $III' = O(1/p^{a_1 - ca_2})$. Thus, $\Delta_{G1}^{(b)} = O(1/p^{a_1 - ca_2})$. Then, we obtain

$$\Delta_{G1} = O_P\left(p^{\frac{7}{2}(1-a_1)}\Big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\Big) + \frac{1}{p^{\frac{7}{2}a_1 - \frac{5}{2} - c}} + \frac{1}{p^{a_1 - ca_2}}\right). \tag{A.12}$$

For Δ_{G3}, we have

$$\Delta_{G3} \leq p^{-1/2}\|\mathbf{\Theta}^{-1/2}\mathbf{\Omega}'\widetilde{\mathbf{V}}_g\widetilde{\mathbf{\Gamma}}_g\widetilde{\mathbf{V}}_g'\mathbf{\Omega}\mathbf{\Theta}^{-1/2}\|_F + p^{-1/2}\|\mathbf{\Theta}^{-1/2}\mathbf{\Omega}'\widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{V}}'\mathbf{\Omega}\mathbf{\Theta}^{-1/2}\|_F =: \Delta_{G3}^{(a)} + \Delta_{G3}^{(b)}.$$

By Lemmas A.2 and A.4, we have

$$\begin{aligned}\|\mathbf{\Omega}'\widetilde{\mathbf{V}}_g\|_F =& \|\mathbf{\Omega}'(\widetilde{\mathbf{V}}_g - \mathbf{V})\|_F = O(\sqrt{p}\|\widetilde{\mathbf{V}}_g - \mathbf{V}\|_{\max}) \\ &= O_P(p^{\frac{11}{2}(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{11}{2}a_1 - \frac{9}{2} - c}).\end{aligned}$$

Since $\|\widetilde{\mathbf{\Gamma}}_g\| = O_P(p^{a_1})$, we have

$$\Delta_{G3}^{(a)} \leq p^{-1/2}\|\mathbf{\Theta}^{-1}\|\|\mathbf{\Omega}'\widetilde{\mathbf{V}}_g\|_F^2\|\widetilde{\mathbf{\Gamma}}_g\| = O_P(p^{\frac{21}{2} - 10a_1}(\log p/T^\alpha + 1/T^{2(1-\alpha)\beta}) + 1/p^{10a_1 - \frac{17}{2} - 2c}).$$

Similarly, $\Delta_{G3}^{(b)} = O_P(1/p^{5a_1 - \frac{7}{2} - 2c})$ because $\|\mathbf{\Omega}'\widetilde{\mathbf{V}}\|_F = O(\sqrt{p}\|\widetilde{\mathbf{V}} - \mathbf{V}\|_{\max}) = O_P(1/p^{3a_1 - 2 - c})$ by Lemma A.2. Then, we obtain

$$\Delta_{G3} = O_P\left(p^{\frac{21}{2} - 10a_1}\Big(\frac{\log p}{T^\alpha} + \frac{1}{T^{2(1-\alpha)\beta}}\Big) + \frac{1}{p^{10a_1 - \frac{17}{2} - 2c}}\right).$$

Similarly, we can show that the terms Δ_{G2}, Δ_{G4}, Δ_{G5} and Δ_{G6} are dominated by Δ_{G1} and Δ_{G3}. Therefore, we have

$$\begin{aligned}\Delta_G = O_P\Big(&p^{\frac{7}{2}(1-a_1)}\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big) + \frac{1}{p^{\frac{7}{2}a_1 - \frac{5}{2} - c}} + \frac{1}{p^{a_1 - ca_2}} \\ &+ p^{\frac{21}{2} - 10a_1}\big(\frac{\log p}{T^\alpha} + \frac{1}{T^{2(1-\alpha)\beta}}\big) + \frac{1}{p^{10a_1 - \frac{17}{2} - 2c}}\Big).\end{aligned} \tag{A.13}$$

Combining the terms Δ_G, Δ_L and Δ_S together, we complete the proof of (3.3). $\square$

A.3 Proof of Theorem 3.2

We provide useful technical lemmas below.

Lemma A.5. *Under Assumptions 3.1–3.2, for $i \le k$, we have*

$$
\begin{aligned}
|\widehat{\delta}_i/\delta_i - 1| &= O_P\left(p^{1-a_1}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta})\right), \\
\|\widehat{v}_i - v_i\|_\infty &= O_P\left(\frac{1}{p^{3(a_1-\frac{1}{2})-c}} + p^{\frac{5}{2}-3a_1}\Big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\Big)\right).
\end{aligned}
$$

Proof. By Lipschitz condition and Assumption 3.2, we have

$$
\|d^{-1}\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}\|_{\max} = O_P(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}). \tag{A.14}
$$

Then, we can obtain the first statement by Weyl's theorem. We have

$$
d^{-1}\widehat{\mathbf{\Sigma}}_h = \mathbf{B}\mathbf{B}' + \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Sigma}_u + (d^{-1}\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}) = \mathbf{B}\mathbf{B}' + \mathbf{\Sigma}_E + (d^{-1}\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}).
$$

We can treat $\mathbf{B}\mathbf{B}'$ as a low rank matrix and the remaining terms as a perturbation matrix. Note that $\|\mathbf{\Sigma}_E\|_\infty = O(p^c)$. By Theorem 1 of Fan et al. (2018b), Lemma A.2, Assumption 3.1 and (A.14), we have

$$
\begin{aligned}
\|\widehat{v}_i - v_i\|_\infty &\le Cp^{2(1-a_1)}\frac{\|\mathbf{\Sigma}_E\|_\infty}{p^{a_1}\sqrt{p}} + Cp^{2(1-a_1)}\frac{\|d^{-1}\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}\|_{\max}}{p^{a_1-1}\sqrt{p}} \\
&= O_P\left(\frac{1}{p^{3(a_1-\frac{1}{2})-c}} + p^{\frac{5}{2}-3a_1}\Big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\Big)\right).
\end{aligned}
$$

$\square$

Lemma A.6. *Under Assumptions 3.1–3.2, for $i \le r_l$, we have*

$$|\widehat{\kappa}_i^l/\kappa_i^l - 1| = O_P\left(p^{\frac{5}{2}(1-a_1)+c(1-a_2)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5a_1}{2}-\frac{3}{2}-2c+ca_2}\right),$$
$$\|\widehat{\eta}_i^l - \eta_i^l\|_\infty = O_P\left(p^{\frac{5}{2}(1-a_1)+c(\frac{5}{2}-3a_2)}(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}) + \frac{1}{p^{\frac{5a_1}{2}-\frac{3}{2}+c(3a_2-\frac{7}{2})}} + \frac{m_p}{p^{3c(a_2-\frac{1}{2})}}\right).$$

Proof. We have

$$\|\mathbf{\Sigma}_E\| \le \|\mathbf{\Lambda}\mathbf{\Lambda}'\| + \|\mathbf{\Sigma}_u\| \le \|\mathbf{\Lambda}\mathbf{\Lambda}'\| + O(m_p) = O(p^{ca_2}).$$

Let $\mathbf{B}\mathbf{B}' = \widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{V}}'$, where $\widetilde{\mathbf{\Gamma}} = \mathrm{diag}(\bar{\delta}_1, \dots, \bar{\delta}_k)$ and their corresponding leading k eigenvectors $\widetilde{\mathbf{V}} = (\bar{v}_1, \dots, \bar{v}_k)$. Also, we let $\mathbf{\Gamma} = \mathrm{diag}(\delta_1, \dots, \delta_k)$ and the corresponding eigenvectors $\mathbf{V} = (v_1, \dots, v_k)$ of covariance matrix $\mathbf{\Sigma}$. Note that $\|\mathbf{B}\|_{\max} = \|\widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}^{1/2}\|_{\max} = O(1)$. By Lemmas A.1-A.2, we have

$$\begin{aligned}\|\mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}} - \widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}^{\frac{1}{2}}\|_{\max} &\le \|\mathbf{B}\widetilde{\mathbf{\Gamma}}^{-\frac{1}{2}}(\mathbf{\Gamma}^{\frac{1}{2}} - \widetilde{\mathbf{\Gamma}}^{\frac{1}{2}})\|_{\max} + \|(\mathbf{V} - \widetilde{\mathbf{V}})\mathbf{\Gamma}^{\frac{1}{2}}\|_{\max} \\ &\le C\frac{\|\mathbf{\Sigma}_E\|}{p^{a_1}} + C\frac{\|\mathbf{\Sigma}_E\|_\infty}{\sqrt{p^{5a_1-3}}} = o(1).\end{aligned}$$

Hence, we have $\|\mathbf{V}\mathbf{\Gamma}^{\frac{1}{2}}\|_{\max} = O(1)$ and $\|\mathbf{V}\|_{\max} = O(1/\sqrt{p^{a_1}})$. By this fact and the results from Lemmas A.1-A.5, we have

$$\begin{aligned}\|\widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{V}}' - \mathbf{V}\mathbf{\Gamma}\mathbf{V}'\|_{\max} &\le \|\widetilde{\mathbf{V}}(\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma})\widetilde{\mathbf{V}}'\|_{\max} + \|(\widetilde{\mathbf{V}} - \mathbf{V})\mathbf{\Gamma}(\widetilde{\mathbf{V}} - \mathbf{V})'\|_{\max} + 2\|\mathbf{V}\mathbf{\Gamma}(\widetilde{\mathbf{V}} - \mathbf{V})'\|_{\max} \\ &= O(p^{-a_1}\|\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\max} + \sqrt{p^{a_1}}\|\widetilde{\mathbf{V}} - \mathbf{V}\|_{\max}) = O(1/p^{\frac{5a_1}{2}-\frac{3}{2}-c}), \\ \|\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{V}\mathbf{\Gamma}\mathbf{V}'\|_{\max} &\le \|\widehat{\mathbf{V}}(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma})\widehat{\mathbf{V}}'\|_{\max} + \|(\widehat{\mathbf{V}} - \mathbf{V})\mathbf{\Gamma}(\widehat{\mathbf{V}} - \mathbf{V})'\|_{\max} + 2\|\mathbf{V}\mathbf{\Gamma}(\widehat{\mathbf{V}} - \mathbf{V})'\|_{\max} \\ &= O_P(p^{-a_1}\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\max} + \sqrt{p^{a_1}}\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max}) \\ &= O_P(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5a_1}{2}-\frac{3}{2}-c}).\end{aligned}$$

Thus, we have

$$\|\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5a_1}{2}-\frac{3}{2}-c}). \tag{A.15}$$

Then, we have

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}_E - \mathbf{\Sigma}_E\|_{\max} &\leq \|\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}\|_{\max} + \|\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max} \\ &= O_P(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^{\alpha}} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5a_1}{2}-\frac{3}{2}-c}). \end{aligned}\tag{A.16}$$

Therefore, the first statement is followed by (A.16) and the Weyl's theorem.

We decompose the sample covariance matrix $\widehat{\mathbf{\Sigma}}_E^l$ for each group l as follows:

$$\widehat{\mathbf{\Sigma}}_E^l = \Lambda^l\Lambda^{l\prime} + \mathbf{\Sigma}_u^l + (\widehat{\mathbf{\Sigma}}_E^l - \mathbf{\Sigma}_E^l).$$

Then, by Theorem 1 of Fan et al. (2018b), Lemma A.2 and (A.16), we have

$$\begin{aligned}\|\widehat{\eta}_i^l - \eta_i^l\|_{\infty} &\leq Cp_l^{2(1-a_2)}\frac{\|\mathbf{\Sigma}_u^l + (\widehat{\mathbf{\Sigma}}_E^l - \mathbf{\Sigma}_E^l)\|_{\infty}}{p_l^{a_2}\sqrt{p_l}} \\ &\leq Cp_l^{2(1-a_2)}\frac{\|\mathbf{\Sigma}_u^l\|_{\infty}}{p_l^{a_2}\sqrt{p_l}} + Cp_l^{2(1-a_2)}\frac{\|\widehat{\mathbf{\Sigma}}_E^l - \mathbf{\Sigma}_E^l\|_{\max}}{p_l^{a_2-1}\sqrt{p_l}} \\ &= O_P\left(p^{\frac{5}{2}(1-a_1)+c(\frac{5}{2}-3a_2)}\left(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}\right) + \frac{1}{p^{\frac{5a_1}{2}-\frac{3}{2}+c(3a_2-\frac{7}{2})}} + \frac{m_p}{p^{3c(a_2-\frac{1}{2})}}\right).\end{aligned}$$

$\square$

Proof of Theorem 3.2. We first consider (3.4). We have

$$\|\widehat{\mathbf{\Phi}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Phi}}' - \mathbf{\Lambda}\mathbf{\Lambda}'\|_{\max} = \max_l \|\widehat{\Phi}^l\widehat{\Psi}^l\widehat{\Phi}^{j\prime} - \Lambda^l\Lambda^{l\prime}\|_{\max}.$$

For each group l, let $\Lambda^l\Lambda^{l\prime} = \widetilde{\Phi}^l\widetilde{\Psi}^l\widetilde{\Phi}^{l\prime}$, where $\widetilde{\Psi}^l = \mathrm{diag}(\bar{\kappa}_1^l, \ldots, \bar{\kappa}_{r_l}^l)$ and the corresponding eigenvectors $\widetilde{\Phi}^l = (\bar{\eta}_1, \ldots, \bar{\eta}_{r_l})$. In addition, let $\Psi^l = \mathrm{diag}(\kappa_1^l, \ldots, \kappa_{r_l}^l)$ and $\Phi^l = (\eta_1, \ldots, \eta_{r_l})$

to be the leading eigenvalues and the corresponding eigenvectors of $\mathbf{\Sigma}_E^l$, respectively. Then, we have

$$\begin{aligned}
\|\Phi^l \Psi^{l\frac{1}{2}} - \widetilde{\Phi}^l \widetilde{\Psi}^{l\frac{1}{2}}\|_{\max} &\leq \|\Lambda^l \widetilde{\Psi}^{l^{-\frac{1}{2}}}(\Psi^{l\frac{1}{2}} - \widetilde{\Psi}^{l\frac{1}{2}})\|_{\max} + \|(\Phi^l - \widetilde{\Phi}^l)\Psi^{l\frac{1}{2}}\|_{\max} \\
&\leq \frac{\|\mathbf{\Sigma}_u^l\|}{p_l^{a_2}} + \frac{\|\mathbf{\Sigma}_u^l\|_\infty}{\sqrt{p_l^{5a_2-3}}} = o(1). \qquad (A.17)
\end{aligned}$$

Since $\|\Lambda^l\|_{\max} = \|\widetilde{\Phi}^l \widetilde{\Psi}^{l\frac{1}{2}}\|_{\max} = O(1)$, $\|\Phi^l \Psi^{l\frac{1}{2}}\|_{\max} = O(1)$ and $\|\Phi^l\|_{\max} = O(1/\sqrt{p_l^{a_2}})$. Using this fact and results from Lemmas A.1, A.2 and A.6, we can show

$$\begin{aligned}
&\|\widetilde{\Phi}^l \widetilde{\Psi}^l \widetilde{\Phi}^{l\prime} - \Phi^l \Psi^l \Phi^{l\prime}\|_{\max} \leq O(p_l^{-a_2}\|\widetilde{\Psi}^l - \Psi^l\|_{\max} + \sqrt{p_l^{a_2}}\|\widetilde{\Phi}^l - \Phi^l\|_{\max}) = O(m_p/\sqrt{p^{c(5a_2-3)}}), \\
&\|\widehat{\Phi}^l \widehat{\Psi}^l \widehat{\Phi}^{l\prime} - \Phi^l \Psi^l \Phi^{l\prime}\|_{\max} \leq O_P(p_l^{-a_2}\|\widehat{\Psi}^l - \Psi^l\|_{\max} + \sqrt{p_l^{a_2}}\|\widehat{\Phi}^l - \Phi^l\|_{\max}) \\
&\qquad = O_P\left(p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big) + \frac{1}{p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})}} + \frac{m_p}{\sqrt{p^{c(5a_2-3)}}}\right).
\end{aligned}$$

By using these rates, we obtain

$$\begin{aligned}
&\|\widehat{\mathbf{\Phi}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Phi}}' - \mathbf{\Lambda}\mathbf{\Lambda}'\|_{\max} \\
&\quad = O_P\left(p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big) + \frac{1}{p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})}} + \frac{m_p}{\sqrt{p^{c(5a_2-3)}}}\right). \qquad (A.18)
\end{aligned}$$

By (A.14), (A.15) and (A.18), we then have

$$\begin{aligned}
\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\|_{\max} &\leq \|d^{-1}\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}\|_{\max} + \|\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max} + \|\widehat{\mathbf{\Phi}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Phi}}' - \mathbf{\Lambda}\mathbf{\Lambda}'\|_{\max} \\
&= O_P\left(p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big) + \frac{1}{p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})}} + \frac{m_p}{\sqrt{p^{c(5a_2-3)}}}\right). \qquad (A.19)
\end{aligned}$$

By definition, $\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}} - \widehat{\boldsymbol{\Sigma}}_u\|_{\max} = \max_{ij} |s_{ij}(\widehat{\sigma}_{ij}) - \widehat{\sigma}_{ij}| \leq \max_{ij} \tau_{ij} = O_P(\tau)$. Then, we have

$$\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}} - \boldsymbol{\Sigma}_u\|_{\max} = O_P(\tau + \omega_{T^\alpha}) = O_P(\omega_{T^\alpha}), \tag{A.20}$$

when τ is chosen as the same order of $\omega_{T^\alpha} = p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5}{2}a_1-\frac{3}{2}+c(\frac{5}{2}a_2-\frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}$. Therefore, by the results of (A.15), (A.18) and (A.20), we have

$$\|\widehat{\boldsymbol{\Sigma}}^{\mathcal{D}} - \boldsymbol{\Sigma}\|_{\max} \leq \|\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max} + \|\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}'\|_{\max} + \|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}} - \boldsymbol{\Sigma}_u\|_{\max} = O_P(\omega_{T^\alpha}).$$

Consider (3.5). Similar to the proofs of Theorem 2.1 in Fan et al. (2011), we can show $\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}} - \boldsymbol{\Sigma}_u\|_2 = O_P(m_p\omega_{T^\alpha}^{1-q})$. In addition, since $\lambda_{\min}(\boldsymbol{\Sigma}_u) > c_1$ and $m_p\omega_{T^\alpha}^{1-q} = o(1)$, the minimum eigenvalue of $\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}}$ is strictly bigger than 0 with probability approaching 1. Then, we have

$$\|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\|_2 \leq \lambda_{\min}(\boldsymbol{\Sigma}_u)^{-1}\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}} - \boldsymbol{\Sigma}_u\|_2\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1} = O_P(m_p\omega_{T^\alpha}^{1-q}). \tag{A.21}$$

Define $\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}} = \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' + \widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}}$. We first show that $\|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_E^{-1}\| = O_P(p^{\frac{c}{2}(1-a_2)}\omega_{T^\alpha} + m_p\omega_{T^\alpha}^{1-q})$. Let $\widehat{\mathbf{J}} = \widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}\widehat{\boldsymbol{\Phi}}'(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}$ and $\widetilde{\mathbf{J}} = \widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}'\boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}$. Using the Sherman-Morrison-Woodbury formula, we have

$$\|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_E^{-1}\| \leq \|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\| + \Delta_{1'},$$

where $\Delta_{1'} = \|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}(\mathbf{I}_r + \widehat{\mathbf{J}})^{-1}\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}\widehat{\boldsymbol{\Phi}}'(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}(\mathbf{I}_r + \widetilde{\mathbf{J}})^{-1}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}'\boldsymbol{\Sigma}_u^{-1}\|$. Then, the right hand side can be bounded by following terms:

$$L_{1'} = \|((\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_u^{-1})\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}(\mathbf{I}_r + \widetilde{\mathbf{J}})^{-1}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}'\boldsymbol{\Sigma}_u^{-1}\|,$$
$$L_{2'} = \|\boldsymbol{\Sigma}_u^{-1}(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}} - \widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}})(\mathbf{I}_r + \widetilde{\mathbf{J}})^{-1}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}'\boldsymbol{\Sigma}_u^{-1}\|,$$

$$L_{3'} = \|\boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}((\mathbf{I}_r+\widehat{\mathbf{J}})^{-1}-(\mathbf{I}_r+\widetilde{\mathbf{J}})^{-1})\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}'\boldsymbol{\Sigma}_u^{-1}\|.$$

By Lemma A.6, $\|\Phi^l\Psi^{l\frac{1}{2}}-\widehat{\Phi}^l\widehat{\Psi}^{l\frac{1}{2}}\|_{\max} \le \|\Lambda^l\widehat{\Psi}^{l^{-\frac{1}{2}}}(\Psi^{l\frac{1}{2}}-\widehat{\Psi}^{l\frac{1}{2}})\|_{\max}+\|(\Phi^l-\widehat{\Phi}^l)\Psi^{l\frac{1}{2}}\|_{\max} = O_P(\omega_{T^\alpha})$, and by (A.17) and (A.21), we then have

$$\begin{aligned}
&\|\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\| \le \max_l \|\widetilde{\Phi}^l\widetilde{\Psi}^{l\frac{1}{2}}\| = O_P(\sqrt{p^{ca_2}}),\\
&\|\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}-\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\| \le \max_l \sqrt{p^c}\|\widehat{\Phi}^l\widehat{\Psi}^{l\frac{1}{2}}-\widetilde{\Phi}^l\widetilde{\Psi}^{l\frac{1}{2}}\|_{\max} = O_P\left(\sqrt{p^c}\omega_{T^\alpha}\right),
\end{aligned}$$

and

$$\begin{aligned}
\|\widehat{\mathbf{J}}-\widetilde{\mathbf{J}}\| &\le \|(\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}\widehat{\boldsymbol{\Phi}}'-\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}')(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1}(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}-\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}})\|\\
&\quad+\|(\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}\widehat{\boldsymbol{\Phi}}'-\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}')(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1}\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\|+\|\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\widetilde{\boldsymbol{\Phi}}'((\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{D}})^{-1}-\boldsymbol{\Sigma}_u^{-1})\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\|\\
&= O_P(p^{\frac{c}{2}(1+a_2)}\omega_{T^\alpha}+p^{ca_2}m_p\omega_{T^\alpha}^{1-q}).
\end{aligned}$$

Since $\lambda_{\min}(\mathbf{I}_r+\widetilde{\mathbf{J}}) \ge \lambda_{\min}(\widetilde{\mathbf{J}}) \ge \lambda_{\min}(\boldsymbol{\Sigma}_u^{-1})\lambda_{\min}^2(\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}) \ge Cp^{ca_2}$, we have $\|(\mathbf{I}_r+\widetilde{\mathbf{J}})^{-1}\| = O_P(1/p^{ca_2})$. Then, $L_{1'} = O_P(m_p\omega_{T^\alpha}^{1-q})$ by (A.21). In addition, $L_{2'} = O_P(p^{-ca_2/2}\|\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}^{\frac{1}{2}}-\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Psi}}^{\frac{1}{2}}\|) = O_P(p^{\frac{c}{2}(1-a_2)}\omega_{T^\alpha})$ and $L_{3'} = O_P(p^{ca_2}\|(\mathbf{I}_r+\widehat{\mathbf{J}})^{-1}-(\mathbf{I}_r+\widetilde{\mathbf{J}})^{-1}\|) = O_P(p^{-ca_2}\|\widehat{\mathbf{J}}-\widetilde{\mathbf{J}}\|) = O_P(p^{\frac{c}{2}(1-a_2)}\omega_{T^\alpha}+m_p\omega_{T^\alpha}^{1-q})$. Thus, we have

$$\Delta_{1'} = O_P(p^{\frac{c}{2}(1-a_2)}\omega_{T^\alpha}+m_p\omega_{T^\alpha}^{1-q}), \tag{A.22}$$

which yields $\|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}})^{-1}-\boldsymbol{\Sigma}_E^{-1}\| = O_P(p^{\frac{c}{2}(1-a_2)}\omega_{T^\alpha}+m_p\omega_{T^\alpha}^{1-q})$.

Let $\widehat{\mathbf{H}} = \widehat{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widehat{\mathbf{V}}'(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}})^{-1}\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}^{\frac{1}{2}}$ and $\widetilde{\mathbf{H}} = \widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'\boldsymbol{\Sigma}_E^{-1}\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}$. Using the Sherman-Morrison-Woodbury formula again, we have

$$\|(\widehat{\boldsymbol{\Sigma}}^{\mathcal{D}})^{-1}-\boldsymbol{\Sigma}^{-1}\| \le \|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}})^{-1}-\boldsymbol{\Sigma}_E^{-1}\|+\Delta_{2'},$$

where $\Delta_{2'} = \|(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}})^{-1}\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}^{\frac{1}{2}}(\mathbf{I}_k+\widehat{\mathbf{H}})^{-1}\widehat{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widehat{\mathbf{V}}'(\widehat{\boldsymbol{\Sigma}}_E^{\mathcal{D}})^{-1}-\boldsymbol{\Sigma}_E^{-1}\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}(\mathbf{I}_k+\widetilde{\mathbf{H}})^{-1}\widetilde{\boldsymbol{\Gamma}}^{\frac{1}{2}}\widetilde{\mathbf{V}}'\boldsymbol{\Sigma}_E^{-1}\|$. By

Weyl's inequality, we have $\lambda_{\min}(\mathbf{\Sigma}_E) > c$ since $\lambda_{\min}(\mathbf{\Sigma}_u) > c$ and $\lambda_{\min}(\mathbf{\Lambda}\mathbf{\Lambda}') = 0$. Hence, $\|\mathbf{\Sigma}_E^{-1}\| = O_P(1)$. By Lemmas A.1-A.5, we have $\|\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}^{\frac{1}{2}} - \widetilde{\mathbf{V}}\widetilde{\mathbf{\Gamma}}^{\frac{1}{2}}\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} - c})$. Similar to the proof of (A.8), we can show $\Delta_{2'} = O_P(m_p\omega_{T^\alpha}^{1-q} + p^{3(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{3a_1-2-c})$. Therefore, we have $\|(\widehat{\mathbf{\Sigma}}^{\mathcal{D}})^{-1} - \mathbf{\Sigma}^{-1}\| = O_P(m_p\omega_{T^\alpha}^{1-q} + p^{\frac{c}{2}(1-a_2)}\omega_{T^\alpha} + p^{3(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{3a_1-2-c})$.

Consider (3.6). We derive the rate of convergence for $\|\widehat{\mathbf{\Sigma}}^{\mathcal{D}} - \mathbf{\Sigma}\|_\Sigma$. The SVD decomposition of $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = (\mathbf{V}_{p\times k} \;\; \mathbf{\Phi}_{p\times r} \;\; \mathbf{\Omega}_{p\times(p-k-r)}) \begin{pmatrix} \mathbf{\Gamma}_{k\times k} & & \\ & \mathbf{\Psi}_{r\times r} & \\ & & \mathbf{\Theta}_{(p-k-r)\times(p-k-r)} \end{pmatrix} \begin{pmatrix} \mathbf{V}' \\ \mathbf{\Phi}' \\ \mathbf{\Omega}' \end{pmatrix}.$$

Note that $\mathbf{\Omega}$ is used to denote the precision matrix in Section 2.2. Moreover, since all the eigenvalues of $\mathbf{\Sigma}$ are strictly bigger than 0, for any maxtrix $\mathbf{A}$, we have $\|\mathbf{A}\|_\Sigma^2 = O_P(p^{-1})\|\mathbf{A}\|_F^2$. Then, we have

$$\begin{aligned}
\|\widehat{\mathbf{\Sigma}}^{\mathcal{D}} - \mathbf{\Sigma}\|_\Sigma \le p^{-1/2}\Big(&\|\mathbf{\Sigma}^{-1/2}(\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\mathbf{\Sigma}^{-1/2}\|_F \\
&+ \|\mathbf{\Sigma}^{-1/2}(\widehat{\mathbf{\Phi}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Phi}}' - \mathbf{\Lambda}\mathbf{\Lambda}')\mathbf{\Sigma}^{-1/2}\|_F + \|\mathbf{\Sigma}^{-1/2}(\widehat{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u)\mathbf{\Sigma}^{-1/2}\|_F\Big) \\
=:\, &\Delta_{G'} + \Delta_{L'} + \Delta_{S'}
\end{aligned}$$

and

$$\Delta_{S'} = O_P(p^{-1/2}\|\widehat{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u\|_F) = O_P(\|\widehat{\mathbf{\Sigma}}_u^{\mathcal{D}} - \mathbf{\Sigma}_u\|_2) = O_P(m_p\omega_{T^\alpha}^{1-q}).$$

We have

$$\begin{aligned}
\Delta_{G'} &= p^{-1/2}\left\|\begin{pmatrix} \mathbf{\Gamma}^{-\frac{1}{2}}\mathbf{V}' \\ \mathbf{\Psi}^{-\frac{1}{2}}\mathbf{\Phi}' \\ \mathbf{\Theta}^{-\frac{1}{2}}\mathbf{\Omega}' \end{pmatrix} (\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\left(\mathbf{V}\mathbf{\Gamma}^{-\frac{1}{2}} \;\; \mathbf{\Phi}\mathbf{\Psi}^{-\frac{1}{2}} \;\; \mathbf{\Omega}\mathbf{\Theta}^{-\frac{1}{2}}\right)\right\|_F \\
&\le p^{-1/2}\big(\|\mathbf{\Gamma}^{-1/2}\mathbf{V}'(\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\mathbf{V}\mathbf{\Gamma}^{-1/2}\|_F + \|\mathbf{\Psi}^{-1/2}\mathbf{\Phi}'(\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\mathbf{\Phi}\mathbf{\Psi}^{-1/2}\|_F
\end{aligned}$$

$$
\begin{aligned}
&+ \|\boldsymbol{\Theta}^{-1/2}\boldsymbol{\Omega}'(\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F + 2\|\boldsymbol{\Gamma}^{-1/2}\mathbf{V}'(\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\boldsymbol{\Phi}\boldsymbol{\Psi}^{-1/2}\|_F \\
&+ 2\|\boldsymbol{\Gamma}^{-1/2}\mathbf{V}'(\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F + 2\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Phi}'(\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F\big) \\
=:\ &\Delta_{G1'} + \Delta_{G2'} + \Delta_{G3'} + 2\Delta_{G4'} + 2\Delta_{G5'} + 2\Delta_{G6'}.
\end{aligned}
$$

In order to find the convergence rate of relative Frobenius norm, we consider the above terms separately. For $\Delta_{G1'}$, we have

$$
\begin{aligned}
\Delta_{G1'} &\leq p^{-1/2}\left(\|\boldsymbol{\Gamma}^{-1/2}\mathbf{V}'(\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}')\mathbf{V}\boldsymbol{\Gamma}^{-1/2}\|_F + \|\boldsymbol{\Gamma}^{-1/2}\mathbf{V}'(\mathbf{V}\boldsymbol{\Gamma}\mathbf{V}' - \mathbf{B}\mathbf{B}')\mathbf{V}\boldsymbol{\Gamma}^{-1/2}\|_F\right) \\
&=: \Delta_{G1'}^{(a)} + \Delta_{G1'}^{(b)}.
\end{aligned}
$$

We bound the two terms separately. We have

$$
\begin{aligned}
\Delta_{G1'}^{(a)} \leq p^{-1/2}\big(&\|\boldsymbol{\Gamma}^{-1/2}(\mathbf{V}'\widehat{\mathbf{V}} - \mathbf{I})\widehat{\boldsymbol{\Gamma}}(\widehat{\mathbf{V}}'\mathbf{V} - \mathbf{I})\boldsymbol{\Gamma}^{-1/2}\|_F + 2\|\boldsymbol{\Gamma}^{-1/2}(\mathbf{V}'\widehat{\mathbf{V}} - \mathbf{I})\widehat{\boldsymbol{\Gamma}}\boldsymbol{\Gamma}^{-1/2}\|_F \\
&+ \|(\boldsymbol{\Gamma}^{-1/2}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{\Gamma}^{-1/2}\|_F\big) =: I + II + III.
\end{aligned}
$$

By Lemma A.5, $\|\mathbf{V}'\widehat{\mathbf{V}} - \mathbf{I}\|_F = \|\mathbf{V}'(\widehat{\mathbf{V}} - \mathbf{V})\|_F \leq \|\widehat{\mathbf{V}} - \mathbf{V}\|_F = O_P(p^{3(1-a_1)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{3a_1-2-c})$. Then, II is of order $O_P(p^{3(1-a_1)-\frac{1}{2}}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{3(a_1-\frac{1}{2})-c})$ and I is of smaller order. In addition, we have $III \leq \|\boldsymbol{\Gamma}^{-1/2}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{\Gamma}^{-1/2}\| = O_P(p^{1-a_1}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}))$ by Lemma A.5. Thus, $\Delta_{G1'}^{(a)} = O_P(p^{1-a_1}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{3(a_1-\frac{1}{2})-c})$. Similarly, we have

$$
\begin{aligned}
\Delta_{G1'}^{(b)} \leq p^{-1/2}\big(&\|\boldsymbol{\Gamma}^{-1/2}(\mathbf{V}'\widetilde{\mathbf{V}} - \mathbf{I})\widetilde{\boldsymbol{\Gamma}}(\widetilde{\mathbf{V}}'\mathbf{V} - \mathbf{I})\boldsymbol{\Gamma}^{-1/2}\|_F + 2\|\boldsymbol{\Gamma}^{-1/2}(\mathbf{V}'\widetilde{\mathbf{V}} - \mathbf{I})\widetilde{\boldsymbol{\Gamma}}\boldsymbol{\Gamma}^{-1/2}\|_F \\
&+ \|(\boldsymbol{\Gamma}^{-1/2}(\widetilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{\Gamma}^{-1/2}\|_F\big) =: I' + II' + III'.
\end{aligned}
$$

By $\sin\theta$ theorem, $\|\mathbf{V}'\widetilde{\mathbf{V}} - \mathbf{I}\| = \|\mathbf{V}'(\widetilde{\mathbf{V}} - \mathbf{V})\| \leq \|\widetilde{\mathbf{V}} - \mathbf{V}\| = O(\|\boldsymbol{\Sigma}_E\|/p^{a_1})$. Then, we have $II' = O(1/p^{a_1-ca_2})$ and I' is of smaller order. By Lemma A.1, we have $III' = O(1/p^{a_1-ca_2})$.

Thus, $\Delta_{G1'}^{(b)} = O(1/p^{a_1 - ca_2})$. Then, we obtain

$$\Delta_{G1'} = O_P\left(p^{1-a_1}(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}) + \frac{1}{p^{3(a_1-\frac{1}{2})-c}} + \frac{1}{p^{a_1-ca_2}}\right). \tag{A.23}$$

For $\Delta_{G3'}$, we have

$$\Delta_{G3'} \le p^{-1/2}\|\boldsymbol{\Theta}^{-1/2}\boldsymbol{\Omega}'\widehat{\mathbf{V}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{V}}'\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F + p^{-1/2}\|\boldsymbol{\Theta}^{-1/2}\boldsymbol{\Omega}'\widetilde{\mathbf{V}}\widetilde{\boldsymbol{\Gamma}}\widetilde{\mathbf{V}}'\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F =: \Delta_{G3'}^{(a)} + \Delta_{G3'}^{(b)}.$$

By Lemma A.5, we have

$$\|\boldsymbol{\Omega}'\widehat{\mathbf{V}}\|_F = \|\boldsymbol{\Omega}'(\widehat{\mathbf{V}}-\mathbf{V})\|_F = O(\sqrt{p}\|\widehat{\mathbf{V}}-\mathbf{V}\|_{\max}) = O_P(p^{3(1-a_1)}(\sqrt{\log p/T^{\alpha}}+1/T^{(1-\alpha)\beta})+1/p^{3a_1-2-c}).$$

Since $\|\widehat{\boldsymbol{\Gamma}}\| = O_P(p^{a_1})$, we have

$$\Delta_{G3'}^{(a)} \le p^{-1/2}\|\boldsymbol{\Theta}^{-1}\|\|\boldsymbol{\Omega}'\widehat{\mathbf{V}}\|_F^2\|\widehat{\boldsymbol{\Gamma}}\| = O_P(p^{11/2-5a_1}(\log p/T^{\alpha} + 1/T^{2(1-\alpha)\beta}) + 1/p^{5a_1-7/2-2c}).$$

Similarly, $\Delta_{G3'}^{(b)} = O_P(1/p^{5a_1-7/2-2c})$ because $\|\boldsymbol{\Omega}'\widetilde{\mathbf{V}}\|_F = O(\sqrt{p}\|\widetilde{\mathbf{V}}-\mathbf{V}\|_{\max}) = O_P(1/p^{3a_1-2-c})$ by Lemma A.2. Then, we obtain

$$\Delta_{G3'} = O_P\left(p^{\frac{11}{2}-5a_1}(\frac{\log p}{T^{\alpha}} + \frac{1}{T^{2(1-\alpha)\beta}}) + \frac{1}{p^{5a_1-\frac{7}{2}-2c}}\right).$$

Similarly, we can show that the terms $\Delta_{G2'}$, $\Delta_{G4'}$, $\Delta_{G5'}$ and $\Delta_{G6'}$ are dominated by $\Delta_{G1'}$ and $\Delta_{G3'}$. Therefore, we have

$$\Delta_{G'} = O_P\left(p^{1-a_1}(\sqrt{\frac{\log p}{T^{\alpha}}} + \frac{1}{T^{(1-\alpha)\beta}}) + \frac{1}{p^{a_1-ca_2}} + p^{\frac{11}{2}-5a_1}(\frac{\log p}{T^{\alpha}} + \frac{1}{T^{2(1-\alpha)\beta}}) + \frac{1}{p^{5a_1-\frac{7}{2}-2c}}\right). \tag{A.24}$$

Similarly, we consider

$$\begin{aligned}
\Delta_{L'} &= p^{-1/2}\left\|\begin{pmatrix}\boldsymbol{\Gamma}^{-\frac{1}{2}}\mathbf{V}'\\ \boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Phi}'\\ \boldsymbol{\Theta}^{-\frac{1}{2}}\boldsymbol{\Omega}'\end{pmatrix}(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\left(\mathbf{V}\boldsymbol{\Gamma}^{-\frac{1}{2}}\quad \boldsymbol{\Phi}\boldsymbol{\Psi}^{-\frac{1}{2}}\quad \boldsymbol{\Omega}\boldsymbol{\Theta}^{-\frac{1}{2}}\right)\right\|_F\\
&\le p^{-1/2}\big(\|\boldsymbol{\Gamma}^{-1/2}\mathbf{V}'(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\mathbf{V}\boldsymbol{\Gamma}^{-1/2}\|_F + \|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Phi}'(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\boldsymbol{\Phi}\boldsymbol{\Psi}^{-1/2}\|_F\\
&\quad + \|\boldsymbol{\Theta}^{-1/2}\boldsymbol{\Omega}'(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F + 2\|\boldsymbol{\Gamma}^{-1/2}\mathbf{V}'(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\boldsymbol{\Phi}\boldsymbol{\Psi}^{-1/2}\|_F\\
&\quad + 2\|\boldsymbol{\Gamma}^{-1/2}\mathbf{V}'(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F + 2\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Phi}'(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\boldsymbol{\Omega}\boldsymbol{\Theta}^{-1/2}\|_F\big)\\
&=: \Delta_{L1'} + \Delta_{L2'} + \Delta_{L3'} + 2\Delta_{L4'} + 2\Delta_{L5'} + 2\Delta_{L6'}.
\end{aligned}$$

For $\Delta_{L2'}$, similar to the proof of (A.24), we have

$$\begin{aligned}
\Delta_{L2'} &\le p^{-1/2}\left(\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Phi}'(\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Phi}}' - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\Phi}')\boldsymbol{\Phi}\boldsymbol{\Psi}^{-1/2}\|_F + \|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Phi}'(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\Phi}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}')\boldsymbol{\Phi}\boldsymbol{\Psi}^{-1/2}\|_F\right)\\
&=: \Delta_{L2'}^{(a)} + \Delta_{L2'}^{(b)}.
\end{aligned}$$

We have

$$\begin{aligned}
\Delta_{L2'}^{(a)} &\le p^{-1/2}\big(\|\boldsymbol{\Psi}^{-1/2}(\boldsymbol{\Phi}'\widehat{\boldsymbol{\Phi}} - \mathbf{I})\widehat{\boldsymbol{\Psi}}(\widehat{\boldsymbol{\Phi}}'\boldsymbol{\Phi} - \mathbf{I})\boldsymbol{\Psi}^{-1/2}\|_F + 2\|\boldsymbol{\Psi}^{-1/2}(\boldsymbol{\Phi}'\widehat{\boldsymbol{\Phi}} - \mathbf{I})\widehat{\boldsymbol{\Psi}}\boldsymbol{\Psi}^{-1/2}\|_F\\
&\quad + \|(\boldsymbol{\Psi}^{-1/2}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1/2}\|_F\big) =: I + II + III.
\end{aligned}$$

By Lemma A.6, we have $\|\widehat{\Phi}^j - \Phi^j\|_F \le \sqrt{p_j r_j}\|\widehat{\Phi}^j - \Phi^j\|_{\max} = O_P\big(p^{\frac{5}{2}(1-a_1)+3c(1-a_2)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5a_1}{2}-\frac{3}{2}+c(3a_2-4)} + m_p/p^{c(3a_2-2)}\big)$. Because $\widehat{\boldsymbol{\Phi}}$ and $\boldsymbol{\Phi}$ are block diagonal matrices, we have

$$\begin{aligned}
\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\|_F^2 &= \sum_{j=1}^{G}\|\widehat{\Phi}^j - \Phi^j\|_F^2\\
&= O_P\left(p^{1-c}\big(p^{5(1-a_1)+6c(1-a_2)}(\frac{\log p}{T^\alpha} + \frac{1}{T^{2(1-\alpha)\beta}}) + \frac{1}{p^{5a_1-3+2c(3a_2-4)}} + \frac{m_p^2}{p^{2c(3a_2-2)}})\right).
\end{aligned}$$

Then, II is of order $O_P(p^{\frac{5}{2}(1-a_1)+c(\frac{5}{2}-3a_2)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5}{2}a_1-\frac{3}{2}+c(3a_2-\frac{7}{2})} + m_p/p^{3c(a_2-\frac{1}{2})})$ and I is of smaller order. Also, $III = O_P(p^{\frac{5}{2}(1-a_1)+c(1-a_2)}(\sqrt{\log p/T^\alpha} + 1/T^{(1-\alpha)\beta})+1/p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2})$ by Lemma A.6. Thus, $\Delta_{L2'}^{(a)} = O_P(p^{\frac{5}{2}(1-a_1)+c(1-a_2)}(\sqrt{\log p/T^\alpha}+ 1/T^{(1-\alpha)\beta}) + 1/p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2} + m_p/p^{3c(a_2-\frac{1}{2})})$. Similarly, we have

$$\Delta_{L2'}^{(b)} \leq p^{-1/2}\big(\|\mathbf{\Psi}^{-1/2}(\mathbf{\Phi}'\widetilde{\mathbf{\Phi}} - \mathbf{I})\widetilde{\mathbf{\Psi}}(\widetilde{\mathbf{\Phi}}'\mathbf{\Phi} - \mathbf{I})\mathbf{\Psi}^{-1/2}\|_F + 2\|\mathbf{\Psi}^{-1/2}(\mathbf{\Phi}'\widetilde{\mathbf{\Phi}} - \mathbf{I})\widetilde{\mathbf{\Psi}}\mathbf{\Psi}^{-1/2}\|_F$$
$$+ \|(\mathbf{\Psi}^{-1/2}(\widetilde{\mathbf{\Psi}} - \mathbf{\Psi})\mathbf{\Psi}^{-1/2}\|_F\big) =: I' + II' + III'.$$

By $\sin\theta$ theorem, $\|\mathbf{\Phi}'\widetilde{\mathbf{\Phi}} - \mathbf{I}\| \leq \|\widetilde{\mathbf{\Phi}} - \mathbf{\Phi}\| \leq \max_j \|\widetilde{\Phi}^j - \Phi^j\| \leq O(m_p/p^{ca_2})$. Then, we have $II' = O(m_p/p^{ca_2})$ and I' is of smaller order. By Lemma A.1, we have $III' = O(m_p/p^{ca_2})$. Thus, $\Delta_{L2'}^{(b)} = O(m_p/p^{ca_2})$. Then, we obtain

$$\Delta_{L2'} = O_P\left(p^{\frac{5}{2}(1-a_1)+c(1-a_2)}\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big) + \frac{1}{p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2}} + \frac{m_p}{p^{ca_2}}\right). \tag{A.25}$$

For $\Delta_{L3'}$, we have

$$\Delta_{L3'} \leq p^{-1/2}\|\mathbf{\Theta}^{-1/2}\mathbf{\Omega}'\widehat{\mathbf{\Phi}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Phi}}'\mathbf{\Omega}\mathbf{\Theta}^{-1/2}\|_F + p^{-1/2}\|\mathbf{\Theta}^{-1/2}\mathbf{\Omega}'\widetilde{\mathbf{\Phi}}\widetilde{\mathbf{\Psi}}\widetilde{\mathbf{\Phi}}'\mathbf{\Omega}\mathbf{\Theta}^{-1/2}\|_F =: \Delta_{L3'}^{(a)} + \Delta_{L3'}^{(b)}.$$

Since $\|\widehat{\mathbf{\Psi}}\| = O_P(p^{ca_2})$, we have

$$\Delta_{L3'}^{(a)} \leq p^{-1/2}\|\mathbf{\Theta}^{-1}\|\|\mathbf{\Omega}'(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})\|_F^2\|\widehat{\mathbf{\Psi}}\|$$
$$= O_P\left(p^{\frac{11}{2}-5a_1+5c(1-a_2)}\big(\frac{\log p}{T^\alpha} + \frac{1}{T^{2(1-\alpha)\beta}}\big) + \frac{1}{p^{5a_1-\frac{7}{2}-c(7-5a_2)}} + \frac{m_p^2}{p^{5ca_2-3c-\frac{1}{2}}}\right).$$

Similarly, by Lemma A.2, $\Delta_{L3'}^{(b)} = O_P(m_p^2/p^{5ca_2-3c-1/2})$ because $\|\widetilde{\Phi}^j - \Phi^j\|_F \leq \sqrt{p_j r_j}\|\widetilde{\Phi}^j - \Phi^j\|_{\max} = O(m_p/p^{c(3a_2-2)})$ and $\|\mathbf{\Omega}'\widetilde{\mathbf{\Phi}}\|_F^2 \leq \|\widetilde{\mathbf{\Phi}}-\mathbf{\Phi}\|_F^2 = \sum_{j=1}^G \|\widetilde{\Phi}^j-\Phi^j\|_F^2 = O(m_p^2/p^{3c(2a_2-1)-1})$. Similarly, we can show $\Delta_{L1'}$, $\Delta_{L4'}$, $\Delta_{L5'}$ and $\Delta_{L6'}$ are dominated by $\Delta_{L2'}$ and $\Delta_{L3'}$. Therefore, we have

$$\Delta_{L'} = O_P\Big(p^{\frac{5}{2}(1-a_1)+c(1-a_2)}\big(\sqrt{\frac{\log p}{T^\alpha}} + \frac{1}{T^{(1-\alpha)\beta}}\big) + \frac{1}{p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2}} + \frac{m_p}{p^{ca_2}}$$
$$+ p^{\frac{11}{2}-5a_1+5c(1-a_2)}\big(\frac{\log p}{T^\alpha} + \frac{1}{T^{2(1-\alpha)\beta}}\big) + \frac{1}{p^{5a_1-\frac{7}{2}-c(7-5a_2)}} + \frac{m_p^2}{p^{5ca_2-3c-\frac{1}{2}}}\Big). \quad \text{(A.26)}$$

Combining the terms $\Delta_{G'}$, $\Delta_{L'}$ and $\Delta_{S'}$ together, we complete the proof of (3.6). $\square$