

# Large Volatility Matrix Analysis Using Global and National Factor Models

Sung Hoon Choi\*      Donggyu Kim<sup>†</sup>

December, 2022

## Abstract

Several large volatility matrix inference procedures have been developed, based on the latent factor model. They often assumed that there are a few of common factors, which can account for volatility dynamics. However, several studies have demonstrated the presence of local factors. In particular, when analyzing the global stock market, we often observe that nation-specific factors explain their own country's volatility dynamics. To account for this, we propose the Double Principal Orthogonal complement Thresholding (Double-POET) method, based on multi-level factor models, and also establish its asymptotic properties. Furthermore, we demonstrate the drawback of using the regular principal orthogonal component thresholding (POET) when the local factor structure exists. We also describe the blessing of dimensionality using Double-POET for local covariance matrix estimation. Finally, we investigate the performance of the Double-POET estimator in an out-of-sample portfolio allocation study using international stocks from 20 financial markets.

**Key words:** High-dimensionality, 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](mailto:sung_hoon.choi@uconn.edu).

<sup>†</sup>Corresponding author. College of Business, Korea Advanced Institute of Science and Technology (KAIST), Seoul, Republic of Korea. Email: [donggyukim@kaist.ac.kr](mailto:donggyukim@kaist.ac.kr).

# 1 Introduction

High dimensional factor analysis and principal component analysis (PCA), which are powerful tools for dimension reduction, have been extensively studied (Bai, 2003; Bernanke et al., 2005; Stock and Watson, 2002). They have various applications in economics and finance, such as forecasting macroeconomic variables and optimal portfolio allocation. Recently, a multi-level factor structure with global and local factors has received increasing attention (Ando and Bai, 2016; Bai and Wang, 2015; Choi et al., 2018; Han, 2021). The global factors impact on all individuals, while the local factors only impact on those in the specific group. The local group can be defined by regional, country, or industry level. In many economic and financial applications, the local or country factors naturally exist. For example, Kose et al. (2003) characterized the comovement of international business cycles on the global, regional, and country levels by imposing a multi-level factor structure on a dynamic factor model. Moench et al. (2013) showed that the local factors play an important role in explaining the U.S. real activities. See also Ando and Bai (2015, 2017); Aruoba et al. (2011); Gregory and Head (1999); Hallin and Liška (2011) for related articles. Hence, when the local factor structure is ignored, the conventional factor analysis might yield misleading results.

Based on the factor models, several large volatility matrix estimation procedures have been developed to account for the strong cross-sectional correlation in the stock market. For example, Fan et al. (2008) studied the impacts of covariance matrix estimation on optimal portfolio allocation and portfolio risk assessment when the factors are observable. In contrast, Fan et al. (2013) considered latent factor models and developed the covariance matrix estimation procedure by PCA and thresholding, which is called the principal orthogonal component thresholding (POET). In this procedure, the unobservable factors can be consistently estimated with a large number of assets. Recent studies, such as Ait-Sahalia and Xiu (2017); Fan et al. (2016, 2018a); Fan and Kim (2019); Jung et al. (2022); Wang and Fan (2017), have also been conducted along this direction. However, when considering the global market, we often observe not only the global common factors but also the nation-specific risk

factors (Kose et al., 2003). That is, the single-level factor model cannot sufficiently explain volatility dynamics.

This paper proposes a novel large volatility matrix estimation procedure that incorporates a global and national factor structure as well as a sparse idiosyncratic volatility matrix. Specifically, we consider the global asset market, and, to account for the nation-specific risk factors, we assume that there are latent multi-level factors, such as the global common factors and national risk factors. Since the national risk is a regional risk, we further assume that the local factor membership is known. Under this latent multi-level factor models, we first use the PCA procedure to capture the latent global factors. Then, after removing the latent global factors, we apply the PCA procedure in each national group to accommodate the latent national factors. Finally, to account for the sparse idiosyncratic volatility matrix, we adopt an adaptive thresholding scheme, which we call this the Double Principal Orthogonal complement Thresholding (Double-POET). We then derive convergence rates for Double-POET and its inverse under different matrix norms. When the local factor membership is unknown, we suggest the regularized spectral clustering method (Amini et al., 2013) to detect the latent local structure. We further discuss the benefit of the proposed Double-POET compared to the regular POET procedure. For example, if we ignore the local factors and treat them as idiosyncratic and employ the POET procedure, the POET estimator can be inconsistent. When estimating the local volatility matrix for each country, Double-POET can estimate global factors better than the regular POET estimator. That is, we can find the blessing of dimensionality. The empirical study supports the theoretical findings.

The rest of the paper is organized as follows. Section 2 sets up the model and proposes the Double-POET estimation procedure. Section 3 presents an asymptotic analysis of the Double-POET estimators. The merits of the proposed method are illustrated by a simulation study in Section 4 and by real data application on portfolio allocation in Section 5. In Section 6, we conclude the study. All proofs are presented in Appendix A.

## 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 a matrix  $\mathbf{A}$ , respectively. In addition, we denote by  $\|\mathbf{A}\|_F$ ,  $\|\mathbf{A}\|_2$  (or  $\|\mathbf{A}\|$  for short),  $\|\mathbf{A}\|_\infty$ , and  $\|\mathbf{A}\|_{\max}$  the Frobenius norm, operator norm,  $l_\infty$ -norm and element-wise norm, which are defined respectively as  $\|\mathbf{A}\|_F = \text{tr}^{1/2}(\mathbf{A}'\mathbf{A})$ ,  $\|\mathbf{A}\|_2 = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$ ,  $\|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$ , 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  $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_n)$  with the diagonal block entries as  $\mathbf{A}_1, \dots, \mathbf{A}_n$ .

### 2.1 Multi-Level Factor Model

We consider the following multi-level factor model:

$$y_{it} = b_i' G_t + \lambda_i^{g_i'} f_t^{g_i} + u_{it}, \quad (2.1)$$

where  $y_{it}$  is the observed data for the  $i$ th individual at time  $t$ , for  $i = 1, \dots, p$ , and  $t = 1, \dots, T$ ;  $G_t$  is a  $k \times 1$  vector of unobservable “global” common factors,  $b_i$  indicates the corresponding factor loadings;  $f_t^{g_i}$  is an  $r_{g_i} \times 1$  vector of unobservable “local” factors that only affect the group  $g_i$ ,  $\lambda_i^{g_i}$  indicates the corresponding factor loadings in each group; and  $u_{it}$  is an idiosyncratic error component. We denote the cluster or group membership as  $g_i \in \{1, \dots, J\}$ . Throughout the paper, we assume that the group membership  $\{g_i\}_{i=1}^p$  is known and the global and local factors are latent. In this paper, we consider the nation-specific local factors; thus, the group membership is the country. In addition, the numbers of global factors and the number of local factors in each group are fixed.

Given the group membership, for  $g_i = j$ , we can stack the observations and denote them as  $y_t \equiv (y_t^1, \dots, y_t^J)'$ , where  $y_t^j = (y_{1t}, \dots, y_{p_j t})'$  and the number of individuals  $p_j$  within group  $j$  such that  $p = \sum_{j=1}^J p_j$ . Let  $F_t = (f_t^1, \dots, f_t^J)'$ , where  $f_t^j$  is an  $r_j \times 1$  vector of local factors and the number of factors  $r_j$  for each group  $j$  such that  $r = \sum_{j=1}^J r_j$ . We define the

$p \times r$  block diagonal matrix as

$$\mathbf{\Lambda} = \text{diag}(\Lambda^1, \Lambda^2, \dots, \Lambda^J),$$

where  $\Lambda^j = (\lambda_{1j}^j, \dots, \lambda_{p_j j}^j)'$  is a  $p_j \times r_j$  matrix of local factor loadings for each  $j$ . Then, the model (2.1) can be written in vector form as follows:

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

where  $\mathbf{B} = (b_1, \dots, b_p)'$  and  $u_t = (u_{1t}, \dots, u_{pt})'$ . In matrix notation, we have

$$\mathbf{Y} = \mathbf{G}\mathbf{B}' + \mathbf{F}\mathbf{\Lambda}' + \mathbf{U}, \quad (2.2)$$

where  $\mathbf{Y} = (y_1, \dots, y_T)'$  is a  $T \times p$  matrix of observed data,  $\mathbf{G} = (G_1, \dots, G_T)'$  is a  $T \times k$  matrix of global factors,  $\mathbf{F} = (F_1, \dots, F_T)'$  is a  $T \times r$  matrix of local factors, and  $\mathbf{U} = (u_1, \dots, u_T)'$  is a  $T \times p$  matrix of idiosyncratic errors that are uncorrelated with global and local factors. Throughout the paper, we further assume that global and local factors are orthogonal each other.

In this paper, the parameter of interest is the  $p \times p$  covariance matrix,  $\mathbf{\Sigma}$ , of  $y_t$  as well as its inverse. Given model (2.1), the covariance matrix can be written as

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

where  $\mathbf{\Sigma}_u$  is a sparse idiosyncratic covariance matrix. We can demonstrate the multi-level factor analysis in the presence of spiked eigenvalues at different levels. Specifically, decomposition (2.3) can be written as

$$\mathbf{\Sigma} = \mathbf{B}\text{cov}(G_t)\mathbf{B}' + \mathbf{\Sigma}_E, \quad (2.4)$$

where  $\Sigma_E = \Lambda \text{cov}(F_t) \Lambda' + \Sigma_u$ . Decomposition (2.4) is a usual single-level factor-based covariance matrix. Thus, based on (2.4), we can apply the regular POET procedure. However, the eigenvalue of the covariance matrix  $\Sigma_E$  diverges, which makes the POET procedure inefficient. In Section 3, we discuss this inefficiency of the regular POET procedure. To further accommodate the local factor structure, we decompose the covariance matrix  $\Sigma_E$  as follows:

$$\Sigma_E^j = \Lambda^j \text{cov}(f_t^j) \Lambda^{j'} + \Sigma_u^j, \quad \text{for } j = 1, \dots, J, \quad (2.5)$$

where the  $j$ th group covariance matrix  $\Sigma_E^j$  is a  $p_j \times p_j$  diagonal block of  $\Sigma_E$ . We assume that the idiosyncratic covariance matrix  $\Sigma_u = (\sigma_{u,ij})_{p \times p}$  is sparse as follows:

$$m_p = \max_{i \leq p} \sum_{j \leq p} |\sigma_{u,ij}|^q, \quad (2.6)$$

for some  $q \in [0, 1)$ , where  $m_p$  diverges slowly, such as  $\log p$ . Intuitively, after removing the global and local factor components, most of pairs are weakly correlated (Bickel and Levina, 2008; Cai and Liu, 2011). Theoretically, since  $\|\Sigma_u\| \leq \|\Sigma_u\|_1 = O(m_p)$ , when  $m_p = o(p_j)$  for all  $j \leq J$ , there are distinguished eigenvalues between the local factor components and the idiosyncratic error components. In light of this, in many applications, the sparsity condition on the factor model residuals has been considered (Boivin and Ng, 2006; Fan et al., 2016). Thus,  $\Sigma_E^j$  is the low-rank plus sparse structure, which has been widely used (Ait-Sahalia and Xiu, 2017; Cai et al., 2013; Candès and Recht, 2009; Fan and Kim, 2019; Fan et al., 2008, 2013; Johnstone and Lu, 2009; Ma, 2013; Negahban and Wainwright, 2011). When considering the U.S. market only, we have the usual single-level factor-based form. Then, based on (2.5), we can apply the regular POET procedure to estimate the local covariance matrix. However, this approach does not use assets outside of the local group, which causes inefficiency. We also discuss this inefficiency in Section 3. To accommodate the multi-level factor structure, we propose a novel large covariance matrix estimation procedure in the following section.

## 2.2 Double-POET Procedure

To make model (2.1) identifiable, without loss of generality, we impose normalization conditions:  $\text{cov}(G_t) = \mathbf{I}_k$  and  $\text{cov}(f_t^j) = \mathbf{I}_{r_j}$ , where  $G_t$  and  $f_t^j$  are uncorrelated with each other; both  $\mathbf{B}'\mathbf{B}$  and  $\Lambda^{j'}\Lambda^j$  for  $j \in \{1, \dots, J\}$  are diagonal matrices. We also assume the pervasiveness conditions, such that (i) the eigenvalues of  $p^{-a_1}\mathbf{B}'\mathbf{B}$  are strictly greater than zero, and (ii) the eigenvalues of  $p_j^{-a_2}\Lambda^{j'}\Lambda^j$  are strictly greater than zero for each  $j$ , where  $a_1, a_2 \in (0, 1]$  are the strengths of global and local factors, respectively. Then, 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  $O(p^{ca_2})$ , which does not grow too fast as  $p^{a_1} \rightarrow \infty$ . We note that  $p^c$  is related to the number of stocks in each country. In addition, all eigenvalues of  $\mathbf{\Sigma}_u$  are bounded. Let  $\mathbf{\Gamma} = \text{diag}(\delta_1, \dots, \delta_k)$  be the leading eigenvalues of  $\mathbf{\Sigma}$  and  $\mathbf{V} = (v_1, \dots, v_k)$  be their corresponding leading eigenvectors.

To accommodate the multi-level factor model (2.1), we propose a non-parametric estimator of  $\mathbf{\Sigma}$  as follows:

1. Let  $\widehat{\delta}_1 \geq \widehat{\delta}_2 \geq \dots \geq \widehat{\delta}_p$  be the ordered eigenvalues of the sample covariance matrix  $\widehat{\mathbf{\Sigma}} = T^{-1} \sum_{t=1}^T (y_t - \bar{y})(y_t - \bar{y})'$  and  $\{\widehat{v}_i\}_{i=1}^p$  be their corresponding eigenvectors. Then, we compute

$$\widehat{\mathbf{\Sigma}}_E = \widehat{\mathbf{\Sigma}} - \widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}',$$

where  $\widehat{\mathbf{\Gamma}} = \text{diag}(\widehat{\delta}_1, \dots, \widehat{\delta}_k)$  and  $\widehat{\mathbf{V}} = (\widehat{v}_1, \dots, \widehat{v}_k)$ .

2. Define  $\widehat{\mathbf{\Sigma}}_E^j$  as each  $p_j \times p_j$  diagonal block of  $\widehat{\mathbf{\Sigma}}_E$ . For the  $j$ th block, let  $\{\widehat{\kappa}_i^j, \widehat{\eta}_i^j\}_{i=1}^{p_j}$  be the eigenvalues and eigenvectors of  $\widehat{\mathbf{\Sigma}}_E^j$  in decreasing order. Then, we compute the principal orthogonal complement  $\widehat{\mathbf{\Sigma}}_u$  as follows:

$$\widehat{\mathbf{\Sigma}}_u = \widehat{\mathbf{\Sigma}} - \widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \widehat{\mathbf{\Phi}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Phi}}',$$

where  $\widehat{\mathbf{\Psi}} = \text{diag}(\widehat{\Psi}^1, \dots, \widehat{\Psi}^J)$  for  $\widehat{\Psi}^j = \text{diag}(\widehat{\kappa}_1^j, \dots, \widehat{\kappa}_{r_j}^j)$ , and the block diagonal matrix

$\widehat{\Phi} = \text{diag}(\widehat{\Phi}^1, \dots, \widehat{\Phi}^J)$  for  $\widehat{\Phi}^j = (\widehat{\eta}_1^j, \dots, \widehat{\eta}_{r_j}^j)$  for  $j = 1, 2, \dots, J$ .

3. Apply the adaptive thresholding method on  $\widehat{\Sigma}_u = (\widehat{\sigma}_{u,ij})_{p \times p}$  following [Bickel and Levina \(2008\)](#) and [Fan et al. \(2013\)](#). Specifically, define  $\widehat{\Sigma}_u^{\mathcal{D}}$  as the thresholded error covariance matrix estimator:

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

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 shrinkage function (e.g., hard or soft thresholding; see [Cai and Liu \(2011\)](#); [Rothman et al. \(2009\)](#)). The thresholding constant  $\tau$  will be determined in [Theorem 3.1](#).

4. The final estimator of  $\Sigma$  is then defined as

$$\widehat{\Sigma}^{\mathcal{D}} = \widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}' + \widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' + \widehat{\Sigma}_u^{\mathcal{D}}. \quad (2.7)$$

The above procedure can be equivalently represented by a least squares method. In particular, the global and local factor matrices and corresponding loading matrices are estimated as follows. To obtain the global factor part, we first solve the following optimization problem:

$$(\widehat{\mathbf{B}}, \widehat{\mathbf{G}}) = \arg \min_{\mathbf{B}, \mathbf{G}} \|\mathbf{Y} - \mathbf{G}\mathbf{B}'\|_F^2, \quad (2.8)$$

subject to the normalization constraints such that

$$\frac{1}{T}\mathbf{G}'\mathbf{G} = \mathbf{I}_k \quad \text{and} \quad \mathbf{B}'\mathbf{B} \text{ is diagonal matrix.}$$

The columns of  $\widehat{\mathbf{G}}/\sqrt{T}$  are the eigenvectors corresponding to the  $k$  largest eigenvalues of the  $T \times T$  matrix  $T^{-1}\mathbf{Y}\mathbf{Y}'$  and  $\widehat{\mathbf{B}} = (\widehat{b}_1, \dots, \widehat{b}_p)' = T^{-1}\mathbf{Y}'\widehat{\mathbf{G}}$ . Given  $\widehat{\mathbf{G}}$  and  $\widehat{\mathbf{B}}$ , let a  $T \times p$  matrix

$\widehat{\mathbf{E}} = \mathbf{Y} - \widehat{\mathbf{G}}\widehat{\mathbf{B}}' \equiv (\widehat{E}^1, \dots, \widehat{E}^J)$ . Then, with  $\widehat{\mathbf{E}}$ , we estimate the local factor part as follows:

$$(\widehat{\Lambda}^j, \widehat{F}^j) = \arg \min_{\Lambda^j, F^j} \|\widehat{E}^j - F^j \Lambda^{j'}\|_F^2, \quad (2.9)$$

subject to the normalization constraints such that

$$\frac{1}{T} F^{j'} F^j = \mathbf{I}_{r_j} \quad \text{and} \quad \Lambda^{j'} \Lambda^j \text{ is diagonal matrix.}$$

Then, we obtain  $\widehat{\mathbf{F}} = (\widehat{F}^1, \dots, \widehat{F}^J)$  and  $\widehat{\Lambda} = \text{diag}(\widehat{\Lambda}^1, \dots, \widehat{\Lambda}^J)$ , where the columns of  $\widehat{F}^j/\sqrt{T}$  are the eigenvectors corresponding to the  $r_j$  largest eigenvalues of the  $T \times T$  matrix  $T^{-1} \widehat{E}^j \widehat{E}^{j'}$  and  $\widehat{\Lambda}^j = (\widehat{\lambda}_1^j, \dots, \widehat{\lambda}_{r_j}^j) = T^{-1} \widehat{E}^{j'} \widehat{F}^j$  for  $j = 1, 2, \dots, J$ . Finally, we apply the above adaptive thresholding method to  $\widehat{\Sigma}_u = T^{-1} \widehat{\mathbf{U}}' \widehat{\mathbf{U}}$ , where  $\widehat{\mathbf{U}} = \mathbf{Y} - \widehat{\mathbf{G}}\widehat{\mathbf{B}}' - \widehat{\mathbf{F}}\widehat{\Lambda}'$ . Similar to the decomposition (2.3), we have the following substitution estimators:

$$\widetilde{\Sigma}^D = \widehat{\mathbf{B}}\widehat{\mathbf{B}}' + \widehat{\Lambda}\widehat{\Lambda}' + \widehat{\Sigma}_u^D. \quad (2.10)$$

Similar to the proofs of Theorem 1 of Fan et al. (2013), we can show that the estimators (2.7) based on PCA and (2.10) based on least squares approaches are equivalent. In particular, by the Eckart-Young theorem, the estimators  $\widehat{\mathbf{B}}$  and  $\widehat{\Lambda}^j$ , after normalization, are the first  $k$  and  $r_j$  empirical eigenvectors of the sample covariance matrix  $\widehat{\Sigma}$  and  $\widehat{\Sigma}_E^j$ , respectively. Then, there exist orthogonal matrices  $\mathbf{H}$  and  $\mathbf{J}$  such that

$$\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{H}'\|_{\max} = O_P(\widetilde{\omega}_T), \quad \|\widehat{\Lambda} - \Lambda\mathbf{J}'\|_{\max} = O_P(\omega_T),$$

where  $\widetilde{\omega}_T$  and  $\omega_T$  are defined in Section 3.1 and Theorem 3.1, respectively. The above rates can be easily obtained using the preliminary results in Appendix A.

In summary, given the knowledge of group membership, we employ PCA on each diagonal block of the remaining components of the sample covariance matrix, after removing the first

$k$  principal components. Then, we apply thresholding on the remaining residual components. We call this procedure the Double Principal Orthogonal complEment Thresholding (Double-POET). Double-POET is the two-step estimation procedure, which makes it possible to estimate global and local factors separately by considering the block structure on the local factors. In contrast, if we employ the single step estimation procedure, such as POET, the local factor structure cannot be explained well. We discuss the theoretical inefficiency of POET in Sections 3.1 and 3.3, and the numerical study illustrated in Section 4 shows that Double-POET outperforms POET.

### 3 Asymptotic Properties

In this section, we establish the asymptotic properties of the Double-POET estimator. To do this, we impose the following technical conditions.

**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  $\Lambda^{j'}\Lambda^j/p_j^{a_2}$  are strictly bigger than zero as  $p, p_j \rightarrow \infty$ , for  $j \in \{1, \dots, J\}$  and  $a_1 \geq ca_2$ . If  $a_1 = ca_2$ , we further assume that  $\lambda_{\min}(\mathbf{B}'\mathbf{B}) - \lambda_{\max}(\Lambda'\Lambda) \geq d_\lambda$  for some positive constant  $d_\lambda$ . In addition, there is a constant  $d > 0$  such that  $\|\mathbf{B}\|_{\max} \leq d$  and  $\|\Lambda\|_{\max} \leq d$ .

(ii) There exists constants  $d_1, d_2 > 0$  such that  $\lambda_{\min}(\Sigma_u) > d_1$  and  $\|\Sigma_u\|_1 \leq d_2 m_p$ .

(iii) The sample covariance matrix  $\widehat{\Sigma}$  satisfies

$$\|\widehat{\Sigma} - \Sigma\|_{\max} = O_P(\sqrt{\log p/T}).$$

**Remark 3.1.** Assumption 3.1(i) ensures that the factors are pervasive. This pervasive assumption is essential to analyze low-rank matrices (Fan et al., 2013, 2018b) and is reasonable

in many financial applications (Bai, 2003; Chamberlain and Rothschild, 1983; Kim and Fan, 2019; Lam and Yao, 2012; Stock and Watson, 2002). For example, under the multi-level factor model, the global factors impact on most of the assets, while the national factors affect only those belonging to each country. This structure of the latent factors implies the pervasive condition, which is related to the incoherence structure (Fan et al., 2018b). To analyze large matrix inferences, we impose the element-wise convergence condition (Assumption 3.1(iii)). Under the sub-Gaussian condition and the mixing time dependency, as considered in Fan et al. (2013), this condition can be easily satisfied (Fan et al., 2018a,b; Vershynin, 2010; Wang and Fan, 2017). We can obtain this condition under the heavy-tailed observations with the condition of bounded fourth moments (Fan et al., 2017, 2018a,b, 2021). Furthermore, when observations are martingales with bounded fourth moments, we can obtain the element-wise convergence condition (Fan and Kim, 2018; Shin et al., 2021).

To measure large matrix estimation errors, we consider the relative Frobenius norm, introduced by Stein and James (1961):

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

Note that the factor  $p^{-1/2}$  performs normalization, such that  $\|\Sigma\|_{\Sigma} = 1$ . Fan et al. (2013) showed that, under this relative Frobenius norm, the POET estimator is consistent as long as  $p = o(T^2)$ , while the sample covariance is consistent only if  $p = o(T)$  in the approximate single-level factor model. In Section 3.1, we will compare the convergence rates of POET and Double-POET in a multi-level factor model.

The following theorem provides the convergence rates under various norms.

**Theorem 3.1.** *Under Assumption 3.1, suppose that  $p_j \asymp p^c$  for each  $j$  and  $m_p = o(p^{c(5a_2-3)/2})$ . Let  $\omega_T = p^{\frac{5}{2}(1-a_1) + \frac{5}{2}c(1-a_2)} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}$  and  $\tau \asymp \omega_T$ . If*

$m_p \omega_T^{1-q} = o(1)$ , we have

$$\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_{\max} = O_P(\omega_T), \quad (3.1)$$

$$\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_2 = O_P(m_p \omega_T^{1-q}), \quad \|(\widehat{\Sigma}_u^{\mathcal{D}})^{-1} - \Sigma_u^{-1}\|_2 = O_P(m_p \omega_T^{1-q}), \quad (3.2)$$

$$\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\max} = O_P(\omega_T), \quad (3.3)$$

$$\|(\widehat{\Sigma}^{\mathcal{D}})^{-1} - \Sigma^{-1}\|_2 = O_P(m_p \omega_T^{1-q}). \quad (3.4)$$

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

$$\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma} = O_P \left( m_p \omega_T^{1-q} + p^{\frac{11}{2} - 5a_1 + 5c(1-a_2)} \frac{\log p}{T} + \frac{1}{p^{5a_1 - \frac{7}{2} - c(7-5a_2)}} + \frac{m_p^2}{p^{5ca_2 - 3c - \frac{1}{2}}} \right). \quad (3.5)$$

**Remark 3.2.** The additional terms  $1/p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})}$  and  $m_p/\sqrt{p^{c(5a_2-3)}}$  in  $\omega_T$  result from the estimation of unknown global factors and national factors. They are negligible when the dimensional  $p$  is high as long as  $0 < c < \frac{5a_1-3}{7-5a_2}$ . In contrast, for the regular POET in the single-level factor model,  $1/\sqrt{p}$  appears in the threshold (Fan et al., 2013). Under the relative Frobenius norm, the upper and lower bounds of  $c$  are required to maintain the consistency of the Double-POET estimator. Specifically, when  $m_p = O(1)$ , the national factor estimation requires  $\frac{1}{2(5a_2-3)} < c < \frac{10a_1-7}{14-10a_2}$ , which is satisfied when both strength levels of global and local factors are sufficiently large.

**Remark 3.3.** The asymptotic theory relies on the relative rates of the number of assets in each group,  $p_j \asymp p^c$ , as well as the number of groups,  $G \asymp p^{1-c}$ . We note that the total number of local factors grows at a rate  $O(p^{1-c})$ . For simplicity, consider the case of strong global and local factors (i.e.,  $a_1 = 1$  and  $a_2 = 1$ ). Hierarchically, to effectively estimate the total local factors, there should be sufficiently large number of assets in each group, but the number of group should not grow much faster than the number of assets in each group to enjoy the blessing of dimensionality. Otherwise, the sum of local factor estimation errors will diverge as  $G$  grows, which causes the curse of dimensionality. Practically, the number

of assets in each country is sufficiently larger than the number of countries. Hence, the Double-POET method can account for the global stock market data.

### 3.1 POET for the Multi-Level Factor Models

In this subsection, we analyze the regular POET of Fan et al. (2013) in the multi-level factor models and compare it with the proposed Double-POET.

The regular POET method only captures the global factors and regards both the local factor structure and idiosyncratic terms as the idiosyncratic part. That is, the sparsity level of  $\Sigma_E = (\varepsilon_{ij})_{p \times p}$  is

$$\mu_p = \max_{i \leq p} \sum_{j \leq p} |\varepsilon_{ij}|^q,$$

for some  $q \in [0, 1)$ . We note that when  $q = 0$ ,  $\mu_p \asymp p^c$ , which corresponds to the maximum number of nonzero elements in each row of  $\Sigma_E$ . Then, the thresholding approach is applied to  $\widehat{\Sigma}_E$  instead of  $\widehat{\Sigma}_u$  to obtain  $\widehat{\Sigma}_E^{\mathcal{T}}$ . Therefore, the regular POET estimator is defined as

$$\widehat{\Sigma}^{\mathcal{T}} = \widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}' + \widehat{\Sigma}_E^{\mathcal{T}}.$$

Similar to the proofs of Fan et al. (2013), we can show that the regular POET yields

$$\|\widehat{\Sigma}_E^{\mathcal{T}} - \Sigma_E\|_2 = O_P(\mu_p \widetilde{\omega}_T^{1-q}), \quad (3.6)$$

$$\|\widehat{\Sigma}^{\mathcal{T}} - \Sigma\|_{\Sigma} = O_P\left(\mu_p \widetilde{\omega}_T^{1-q} + p^{\frac{11}{2} - 5a_1} \frac{\log p}{T} + \frac{1}{p^{5a_1 - \frac{7}{2} - 2c}}\right), \quad (3.7)$$

where  $\widetilde{\omega}_T = p^{\frac{5}{2}(1-a_1)} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} - c}$ . We compare the convergence rates of the regular POET and proposed Double-POET. For example, under the relative Frobenius norm, when  $q = 0$ ,  $m_p = O(1)$ ,  $a_1 = 1$ , and  $a_2 = 1$ , we have

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

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

The number of assets within a group,  $p_j \asymp p^c$ , for some constant  $c > 0$  plays a crucial role in a convergence rate. Theorem 3.1 shows that  $\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma}$  can be convergent as long as  $p = o(T^2)$  and  $\frac{1}{4} < c < \frac{3}{4}$ . In contrast, the rate of the regular POET estimator,  $\|\widehat{\Sigma}^{\mathcal{T}} - \Sigma\|_{\Sigma}$ , does not converge if  $c > \frac{1}{2}$  or  $p^{\alpha} > T$  with  $\alpha = \min\{\frac{1}{2}, 2c\}$ . In other words, the regular POET requires the number of assets in each country to be small enough to avoid the curse of dimensionality. However, the number of assets in each country is larger than the number of countries; thus, it is more reasonable to assume  $c > \frac{1}{2}$ . Therefore, under the global and national factor models, the regular POET does not provide a consistent estimator in terms of the relative Frobenius norm. Furthermore, when the global factors are weak (i.e.,  $a_1 < 1$ ) and the local factors are strong (i.e.,  $a_2 = 1$ ), Double-POET is equal to or better than POET in terms of the convergence rate under the relative Frobenius norm.

## 3.2 Orthogonality between the Global and Local Factor Loadings

When the signal of local factors is strong, we often consider the local factors as the global weak factors. Then, we can apply the regular POET method with the global and local factors. Theoretically, to identify the latent factors, we additionally need to impose an orthogonality condition between the global and local factor loadings,  $\mathbf{B}$  and  $\mathbf{\Lambda}$ . In this section, under this orthogonality condition, we investigate asymptotic behaviors of the Double-POET procedure and compare POET with Double-POET.

Under the orthogonality condition, we first obtain the following modified theoretical results for Double-POET.

**Theorem 3.2.** *Suppose that  $\mathbf{B}$  and  $\mathbf{\Lambda}$  are orthogonal each other. Under Assumption 3.1, suppose that  $p_j \asymp p^c$  for each  $j$  and  $m_p = o(\min\{p^{(5a_1-3)/2}, p^{c(5a_2-3)/2}\})$ . Let  $\omega_{T,2} = p^{\frac{5}{2}(1-a_1) + \frac{5}{2}c(1-a_2)} \sqrt{\log p/T} + m_p/\sqrt{p^{c(5a_2-3)}} + m_p/\sqrt{p^{5a_1-3-5c(1-a_2)}}$  and  $\tau \asymp \omega_{T,2}$ . If  $m_p \omega_{T,2}^{1-q} =$*

$o(1)$ , we have

$$\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_{\max} = O_P(\omega_{T,2}), \quad (3.8)$$

$$\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_2 = O_P(m_p \omega_{T,2}^{1-q}), \quad \|(\widehat{\Sigma}_u^{\mathcal{D}})^{-1} - \Sigma_u^{-1}\|_2 = O_P(m_p \omega_{T,2}^{1-q}), \quad (3.9)$$

$$\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\max} = O_P(\omega_{T,2}), \quad (3.10)$$

$$\|(\widehat{\Sigma}^{\mathcal{D}})^{-1} - \Sigma^{-1}\|_2 = O_P(m_p \omega_{T,2}^{1-q}). \quad (3.11)$$

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

$$\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma} = O_P \left( m_p \omega_{T,2}^{1-q} + p^{\frac{11}{2} - 5a_1 + 5c(1-a_2)} \frac{\log p}{T} + \frac{m_p^2}{p^{5a_1 - \frac{7}{2} - 5c(1-a_2)}} + \frac{m_p^2}{p^{c(5a_2-3) - \frac{1}{2}}} \right). \quad (3.12)$$

**Remark 3.4.** The orthogonality condition helps identify the latent global and local factors by reducing perturbation terms. Thus, compared to the results of Theorem 3.1, Theorem 3.2 shows the faster convergence rate, and we can relax the upper bound for  $c$ . For example, the additional terms  $m_p/\sqrt{p^{c(5a_2-3)}}$  and  $m_p/\sqrt{p^{5a_1-3-5c(1-a_2)}}$  in  $\omega_{T,2}$  are negligible when  $0 < c < \frac{5a_1-3}{5(1-a_2)}$  and  $c \leq \frac{a_1}{a_2}$  as  $p \rightarrow \infty$ . In addition, when  $m_p = O(1)$ ,  $\frac{1}{2(5a_2-3)} < c < \frac{a_1-0.7}{1-a_2}$  and  $c \leq \frac{a_1}{a_2}$  are required to make  $\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma}$  converge. Therefore, the Double-POET method does not require the upper bound for  $c$  when the global and local factors are strong. For example, the number of group,  $J$ , can be fixed (i.e.,  $c = 1$ ) as long as  $a_1 \geq a_2$ .

We now consider the POET estimator that regards both the global and local factors as the global factors under the orthogonality condition. We call this the POET2 estimator. The estimator is then defined as follows:

$$\widehat{\Sigma}_2^{\mathcal{T}} = \widehat{\mathbf{V}}_2 \widehat{\Gamma}_2 \widehat{\mathbf{V}}_2' + \widehat{\Sigma}_{u,2}^{\mathcal{T}},$$

where  $\widehat{\Gamma}_2 = \text{diag}(\widehat{\delta}_1, \dots, \widehat{\delta}_{k+r})$  and  $\widehat{\mathbf{V}}_2 = (\widehat{v}_1, \dots, \widehat{v}_{k+r})$ . Then, when  $a_1 = a_2 = 1$ , the POET2

estimator yields

$$\|\widehat{\Sigma}_{u,2}^{\mathcal{T}} - \Sigma_u\|_2 = O_P(m_p \widetilde{\omega}_{T,2}^{1-q}), \quad (3.13)$$

$$\|\widehat{\Sigma}_2^{\mathcal{T}} - \Sigma\|_{\Sigma} = O_P\left(m_p \widetilde{\omega}_{T,2}^{1-q} + p^{15(1-c)+\frac{1}{2}} \frac{\log p}{T} + \frac{m_p^2}{p^{15c-\frac{27}{2}}}\right), \quad (3.14)$$

where  $\widetilde{\omega}_{T,2} = p^{7(1-c)} \sqrt{\log p/T} + m_p/p^{7c-6}$ . We note that the additional term  $m_p/p^{7c-6}$  in  $\widetilde{\omega}_{T,2}$  can be negligible only when  $c > \frac{6}{7}$  as  $p$  increases. This is because the POET2 estimator includes noises on the off-diagonal blocks of the local factor component. Therefore, it requires the number of groups to be sufficiently small. Otherwise, the POET2 estimator does not perform well. That is, since the local factors are considered as the weak factors of the global factor, the local factor should have enough signals.

We compare the convergence rates of the POET2 and Double-POET estimators under the orthogonality condition. When  $q = 0$ ,  $m_p = O(1)$ ,  $a_1 = 1$ , and  $a_2 = 1$ , we have

$$\begin{aligned} \|\widehat{\Sigma}_2^{\mathcal{T}} - \Sigma\|_{\Sigma} &= O_P\left(p^{7(1-c)} \sqrt{\frac{\log p}{T}} + \frac{1}{p^{7c-6}} + p^{15(1-c)+\frac{1}{2}} \frac{\log p}{T} + \frac{1}{p^{15c-\frac{27}{2}}}\right), \\ \|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma} &= O_P\left(\sqrt{\frac{\log p}{T}} + \frac{1}{p^c} + \frac{\sqrt{p} \log p}{T} + \frac{1}{p^{2c-\frac{1}{2}}}\right). \end{aligned}$$

When  $c < 1$ , the convergence rate of Double-POET is faster than that of POET2. Furthermore,  $\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma}$  can be consistent as long as  $p = o(T^2)$  and  $c > \frac{1}{4}$ , while  $\|\widehat{\Sigma}_2^{\mathcal{T}} - \Sigma\|_{\Sigma}$  can be consistent when  $p^{15(1-c)+\frac{1}{2}} = o(T)$  and  $c > \frac{9}{10}$ . This implies that POET2 performs well only for the multi-level factor model with a small number of groups (i.e., weak factors with enough signals). We also note that when the local factor has the same signal as the global factor ( $c = 1$ ), the POET2 and Double-POET procedures have the same convergence rate.

### 3.3 Blessing of Dimensionality

In this section, we demonstrate the blessing of dimensionality for country-wise covariance matrix estimation by using the proposed Double-POET procedure. For the country  $j$ , we

have

$$y_t^j = B^j G_t + \Lambda^j f_t^j + u_t^j,$$

where  $B^j = (b_1^j, \dots, b_{p_j}^j)'$ ,  $\Lambda^j = (\lambda_1^j, \dots, \lambda_{p_j}^j)'$ , and  $u_t^j = (u_{1t}, \dots, u_{p_j t})'$ . Then, the covariance matrix of the  $j$ th country can be written as

$$\Sigma^j = B^j \text{cov}(G_t) B^{j'} + \Lambda^j \text{cov}(f_t^j) \Lambda^{j'} + \Sigma_u^j.$$

Then, the sparsity level of  $\Sigma_u^j$  is

$$m_{p_j} := \max_{\substack{i \leq p_j \\ j \leq p_j}} |\sigma_{u,ij}|^q, \text{ for some } q \in [0, 1).$$

To estimate the  $j$ th country's covariance matrix  $\Sigma^j$ , assuming that the common factors include both global and national factors with the orthogonality condition between their factor loadings, the regular POET method can be applied with  $k + r_j$  principal components using the  $j$ th country's stock market data (i.e.,  $T \times p_j$  observation matrix), denoted by  $\widehat{\Sigma}^{j,T}$ . As long as  $p_j \rightarrow \infty$ , the asymptotic results of Fan et al. (2013) can be directly applied by replacing  $p$  by  $p_j \asymp p^c$  for  $c \in (0, 1]$ . In contrast, we suggest the Double-POET method by extracting the  $j$ th diagonal block of  $\widehat{\Sigma}^D$ , denoted by  $(\widehat{\Sigma}^D)^j$ . Then, when estimating the global factor component, Double-POET uses more data, which might result in a more accurate global factor estimator. Specifically, we have the following convergence rates of Double-POET estimator for the local covariance matrix.

**Theorem 3.3.** *Under the assumptions of Theorem 3.2, the Double-POET estimator satisfies, if  $m_{p_j} \omega_{T,2}^{1-q} = o(1)$ ,*

$$\|(\widehat{\Sigma}^D)^j - \Sigma^j\|_{\max} = O_P(\omega_{T,2}), \quad (3.15)$$

$$\|((\widehat{\Sigma}^D)^j)^{-1} - (\Sigma^j)^{-1}\|_2 = O_P(m_{p_j} \omega_{T,2}^{1-q}). \quad (3.16)$$

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

$$\|(\widehat{\Sigma}^{\mathcal{D}})^j - \Sigma^j\|_{\Sigma^j} = O_P\left(m_{p_j}\omega_{T,2}^{1-q} + p^{5(1-a_1)+c(\frac{11}{2}-5a_2)}\frac{\log p}{T} + \frac{m_p^2}{p^{5a_1-3+c(5a_2-\frac{11}{2})}} + \frac{m_p^2}{p^{c(5a_2-\frac{7}{2})}}\right), \quad (3.17)$$

where  $\|A\|_{\Sigma^j} = p_j^{-1/2}\|\Sigma^{j-1/2}A\Sigma^{j-1/2}\|_F$  is the relative Frobenius norm.

**Remark 3.5.** We compare the rates of convergence between the Double-POET and POET estimators as follows. Following Fan et al. (2018a), under the assumptions of Theorem 3.3, the regular POET estimator satisfies the following conditions, with  $\tau \asymp \ddot{\omega}_T$ , if  $m_{p_j}\ddot{\omega}_T = o(1)$ ,  $a_1 = 1$ , and  $a_2 = 1$ :

$$\begin{aligned} \|\widehat{\Sigma}^{j,\mathcal{T}} - \Sigma^j\|_{\max} &= O_P(\ddot{\omega}_T), \\ \left\|(\widehat{\Sigma}^{j,\mathcal{T}})^{-1} - (\Sigma^j)^{-1}\right\|_2 &= O_P(m_{p_j}\ddot{\omega}_T^{1-q}), \\ \|\widehat{\Sigma}^{j,\mathcal{T}} - \Sigma^j\|_{\Sigma^j} &= O_P\left(m_{p_j}\ddot{\omega}_T^{1-q} + \frac{\sqrt{p^c}\log p}{T}\right), \end{aligned}$$

where  $\ddot{\omega}_T = \sqrt{\log p/T} + m_{p_j}/p^c$ . The overall rates of convergence under each norm are the same. This might be because the convergence rate is dominated by the national factor estimator, which is estimated by the Double-POET and POET procedures, using the same amount of the data. However, when we focus on the global factor components, the Double-POET procedure indeed can have a faster rate of convergence than POET. For example, when  $a_1 = 1$  and  $a_2 = 1$ , under the max norm, the convergence rates of Double-POET and POET for  $B^j B^{j'}$  are  $O_P(\sqrt{\log p/T} + m_p/p)$  and  $O_P(\sqrt{\log p/T} + m_p/p^c)$ , respectively. In addition, in terms of the relative Frobenius norm, the convergence rates of Double-POET and POET for  $B^j B^{j'}$  are  $O_P(\sqrt{\log p/T} + \sqrt{p^c}\log p/T + m_p/p + m_p^2/p^{2-c/2})$  and  $O_P(1/\sqrt{T} + \sqrt{p^c}\log p/T + m_p^2/p^{3c/2})$ , respectively. Thus, the global factor estimator of Double-POET has a faster convergence rates than that of POET when  $c < 1$  (under the max norm) or  $c < \frac{2}{3}$  when  $m_p = O(1)$  (under the relative Frobenius norm). Furthermore, the numerical study

presented in Section 4 shows that Double-POET outperforms POET, especially for small  $p_j$ . This implies that Double-POET enjoys the blessing of dimensionality by incorporating other countries' observations. That is, the proposed Double-POET model not only presents a way to investigate the global stock market, but also shows the benefits of incorporating financial big data.

### 3.4 Determination of the Number of Factors

To implement Double-POET, we need to determine the number of factors. In this section, we describe a data-driven method for determining the number of global and national factors.

We suggest a modified version of the eigenvalue ratio method proposed by [Ahn and Horenstein \(2013\)](#) as follows. We first consider a model selection problem between the multi-level factor model and the single-level factor model, and then propose estimators for the number of factors. Let  $\hat{\delta}_m$  be the  $m$ th largest eigenvalue of the sample covariance matrix and  $\text{ER}(m) = \hat{\delta}_m / \hat{\delta}_{m+1}$  be the  $m$ th eigenvalue ratio. Under the multi-level factor model (i.e.,  $k > 0$  and  $r > 0$ ), the first  $k$  eigenvalues of the sample covariance matrix are asymptotically determined by the eigenvalues of  $\mathbf{B}\text{cov}(G_t)\mathbf{B}'$ , the next  $r = \sum_{j=1}^J r_j$  eigenvalues by the eigenvalues of  $\mathbf{\Lambda}\text{cov}(F_t)\mathbf{\Lambda}'$ , and the other eigenvalues by those of the idiosyncratic covariance matrix. Accordingly, when  $a_1 = 1$  and  $a_2 = 1$ ,  $\text{ER}(m) = O_P(1)$  for  $m \neq k$  and  $m \neq k + r$ ,  $\text{ER}(k) = O_P(p^{1-c})$ , and  $\text{ER}(k + r) = O_P(p^c)$ . This implies that there are two diverging eigenvalue ratios when  $0 < c < 1$ , while all other ratios of two adjacent eigenvalues are asymptotically bounded. In contrast, under the single-level factor model, there exists only one diverging eigenvalue ratio. We define

$$\hat{k}_1 = \max_{1 \leq m \leq k_{\max}} \text{ER}(m) \quad \text{and} \quad \hat{k}_2 = \max_{1 \leq m \neq \hat{k}_1 \leq k_{\max}} \text{ER}(m),$$

for a prespecified  $k_{\max} < \min(p, T)$ . Let  $\varphi_p$  be the tuning parameter, which grows slowly. In practice, we set  $\varphi_p = d \log p$  for a positive constant  $d$ . We then select the single-level factor

model when  $\text{ER}(\hat{k}_1) > \varphi_p$  and  $\text{ER}(\hat{k}_2) \leq \varphi_p$ , and one can apply the regular POET with  $\hat{k}_1$  factors. When  $\text{ER}(\hat{k}_2) > \varphi_p$ , we select the multi-level factor model and estimate the number of global factors  $k$  by  $\hat{k} = \min\{\hat{k}_1, \hat{k}_2\}$ . Under the following technical conditions, we can show the consistency of  $\hat{k}$ .

**Assumption 3.2.**

- (i) There exist  $T \times T$  and  $p \times p$  positive semidefinite matrices  $R_T$  and  $G_p$  such that  $\mathbf{U} = R_T^{1/2} \Upsilon G_p^{1/2}$ , where the idiosyncratic observation matrix  $\mathbf{U}$  is defined in (2.2), and  $\Upsilon' = [v_{it}]_{p \times T}$ .
- (ii)  $v_{it}$  indicates i.i.d. random variables with uniformly bounded moments up to the fourth order.
- (iii) There are generic positive constants  $d_1, d_2 > 0$  such that  $\lambda_{\max}(R_T) < d_1$ ,  $\lambda_{\max}(G_p) < d_1$ , and  $\lambda_{\min}(R_T) > d_2$ .
- (iv) Let  $y^* = \lim_{m \rightarrow \infty} m/p$ , where  $m = \min(p, T)$ . Then, there exists a real number  $d^* \in (1 - y^*, 1]$  such that the  $(d^*p)$ th largest eigenvalue of  $G_p$  is strictly greater than zero for all  $p$ .

Then, we have the following theorem.

**Theorem 3.4.** Suppose Assumptions 3.1–3.2 hold, and in addition suppose that  $m_p = o(p^{ca_2})$ ,  $c < \frac{5a_1 - 3}{4 - 2a_2}$ , and  $\max\{p^{1-ca_2}, p^{5(1-a_1)+2c(1-a_2)} \log p\} = o(T)$ .

- (i) Given a tuning parameter  $\varphi_p < \min\{p^{a_1-ca_2}, \frac{m}{p^{1-ca_2}}\}$ , the proposed model selection is consistent.
- (ii) When  $k > 0$  and  $r > 0$ , there exists a constant  $d \in (0, 1]$  such that  $\lim_{m \rightarrow \infty} \Pr(\hat{k} = k) = 1$ , for any  $k_{\max} \in (k + r, dm - 2(k + r) - 1]$ .

Given the consistently estimated number of global factors, we can apply the existing methods to consistently estimate the number of local factors  $r_j$  (Alessi et al., 2010; Bai and

Ng, 2002; Choi et al., 2018; Giglio and Xiu, 2021; Onatski, 2010; Trapani, 2018). Throughout the paper, we use the eigenvalue ratio method of Ahn and Horenstein (2013) as follows: for each group  $j$ , let  $\widehat{\kappa}_r^j$  be the  $r$ th largest eigenvalues of the  $p_j \times p_j$  matrix  $\widehat{\Sigma}_E^j$  in the Section 2.2. Then,  $r_j$  can be estimated by

$$\hat{r}_j = \max_{1 \leq r \leq r_{j,\max}} \frac{\widehat{\kappa}_r^j}{\widehat{\kappa}_{r+1}^j},$$

for a predetermined  $r_{j,\max}$ .

### 3.5 Unknown Local Group Membership

In this paper, we assume that local factors are governed by the national regional risk factors, which provides the membership of the local factors. However, in practice, the membership of the local factors is unknown. In this section, we discuss how to use the proposed Double-POET procedure for the unknown local factor membership.

The Double-POET procedure is working as long as the membership for the local factor is known. Thus, to harness the Double-POET, we need to classify assets and find the latent local structure. As discussed in the previous sections, we can estimate the global factor even if the local factor structure is unknown. Thus, we can consistently estimate  $\Sigma_E$  defined in (2.4), which is the remaining covariance matrix after subtracting the global factor part. Under some local factor structure,  $\Sigma_E$  is a form of a block diagonal matrix and its block membership represents the local factor membership. Thus, we can detect the group membership, based on  $\Sigma_E$ . Specifically, to adjust the scale problem of the covariance matrix, we calculate the correlation matrix of  $\Sigma_E$ . Since the sign of the correlations does not contain the membership information, we use the absolute values of the correlations. We denote this matrix by  $\mathbf{L}$ . We consider  $\mathbf{L}$  as the adjacency matrix of the local factor network. Based on the adjacency matrix, many models and methodologies have been developed to detect and identify group memberships. Examples include RatioCut (Hagen and Kahng, 1992), Neut (Shi and Malik,

2000), spectral clustering method (Lei and Rinaldo, 2013; McSherry, 2001; Rohe et al., 2011), regularized spectral clustering (Amini et al., 2013), semi-definite programming (Cai and Li, 2015; Hajek et al., 2015), Newman-Girvan Modularity (Girvan and Newman, 2002), and maximum likelihood estimation (Amini et al., 2013; Bickel and Chen, 2009). To detect the membership matrix, we employ the regularized spectral clustering (RSC) (Amini et al., 2013).

The regularized spectral clustering (RSC) is based on the regularized row and column normalized adjacency matrix (or the regularized graph Laplacian) (Chaudhuri et al., 2012; Qin and Rohe, 2013),

$$\tilde{\mathbf{L}}_{deg,a} = \mathbf{D}_a^{-1/2} \mathbf{L} \mathbf{D}_a^{-1/2},$$

where the degree matrix is denoted by  $\mathbf{D} = \text{diag}(\hat{d}_1, \dots, \hat{d}_p)$  with  $\hat{d}_i = \sum_{j=1}^p L_{ij}$ ,  $\mathbf{D}_a = \mathbf{D} + a\mathbf{I}$ ,  $\mathbf{I}$  denotes the identity matrix, and  $a \geq 0$  is the regularization parameter. For the numerical study, we use the average node degree as the regularization parameter  $a$ . Then, calculate the eigenvector matrix corresponding to the  $K$  largest eigenvalue of  $\tilde{\mathbf{L}}_{deg,a}$ . Using the eigenvector matrix, we apply the k-means clustering procedure and identify the local factor groups. Unfortunately, we cannot observe  $\Sigma_E$ , and so we estimate it using the POET procedure. Then, using the plug-in method, we estimate the regularized row and column normalized adjacency matrix. Under some regularity condition, we can show that the ratio of the misclassification goes to zero (Joseph et al., 2016; Qin and Rohe, 2013). With this estimated membership, we can apply the proposed Double-POET.

## 4 Simulation Study

In this section, we conducted simulations to examine the finite sample performance of the proposed Double-POET method. We considered the following multi-level factor model:

$$y_{it} = \sum_{l=1}^k b_{il} G_{tl} + \sum_{s=1}^{r_j} \lambda_{is}^j f_{ts}^j + u_{it},$$

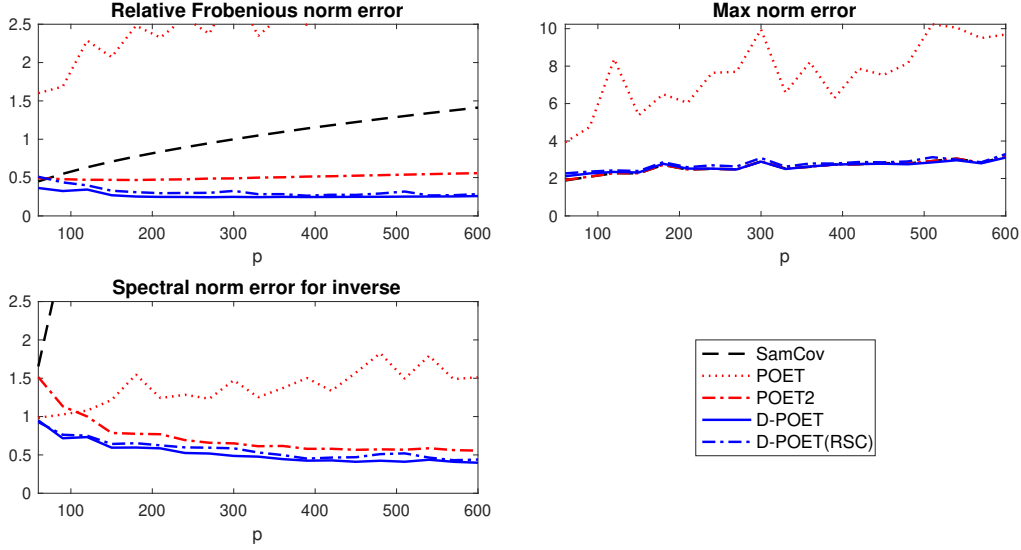


Figure 1: Averages of  $\|\widehat{\Sigma} - \Sigma\|_{\Sigma}$ ,  $\|\widehat{\Sigma} - \Sigma\|_{\max}$  and  $\|(\widehat{\Sigma})^{-1} - \Sigma^{-1}\|$  for Double-POET, Double-POET(RSC), POET, POET2 and the sample covariance matrix against  $p$  with a fixed  $J = 10$ .

where the global factors and national factors,  $G_{it}$  and  $f_{ts}^j$ , respectively, were all drawn from  $\mathcal{N}(0, 1)$ . The global factor loadings  $\{b_{i1}, \dots, b_{ik}\}_{i \leq p}$  were drawn from  $\mathcal{N}(\mu_B, I_k)$ , where each element of  $\mu_B$  is i.i.d.  $\text{Uniform}(-0.5, 0.5)$ ; for each  $j$ , the local factor loadings  $\{\lambda_{i1}^j, \dots, \lambda_{ir_j}^j\}_{i \leq p_j}$  were drawn from  $\mathcal{N}(\mu_{\lambda^j}, I_{r_j})$ , where each element of  $\mu_{\lambda^j}$  is i.i.d.  $\text{Uniform}(-0.3, 0.3)$ .

We generated the idiosyncratic errors as follows. Let  $D = \text{diag}(d_1^2, \dots, d_p^2)$ , where each  $\{d_i\}$  was generated independently from Gamma  $(\alpha, \beta)$  with  $\alpha = \beta = 100$ . We set  $s = (s_1, \dots, s_p)'$  to be a sparse vector, where each  $s_i$  was drawn from  $\mathcal{N}(0, 1)$  with probability  $\frac{m}{\sqrt{p} \log p}$ , and  $s_i = 0$  otherwise. Then, we set a sparse error covariance matrix as  $\Sigma_u = D + ss' - \text{diag}\{s_1^2, \dots, s_p^2\}$ . In the simulation, we generated  $\Sigma_u$  until it is positive definite. Note that varying  $m > 0$  enables us to control the sparsity level, and we chose  $m = 0.3$ . Finally, we generated  $\{u_t\}_{t \leq T}$  from i.i.d.  $\mathcal{N}(0, \Sigma_u)$ .

In this simulation study, we chose the number of periods  $T = 300$  and the numbers of factors as  $k = 3$  and  $r_j = 2$  for each  $j$ . Then, we considered two cases: (i) increasing  $p$  from 60 to 600 in increments of 30 with a fixed  $J = 10$  (i.e., each  $p_j = p/10$ ), and (ii) increasing  $J$  from 2 to 20 with a fixed  $p_j = 30$ . Each simulation is replicated 200 times.

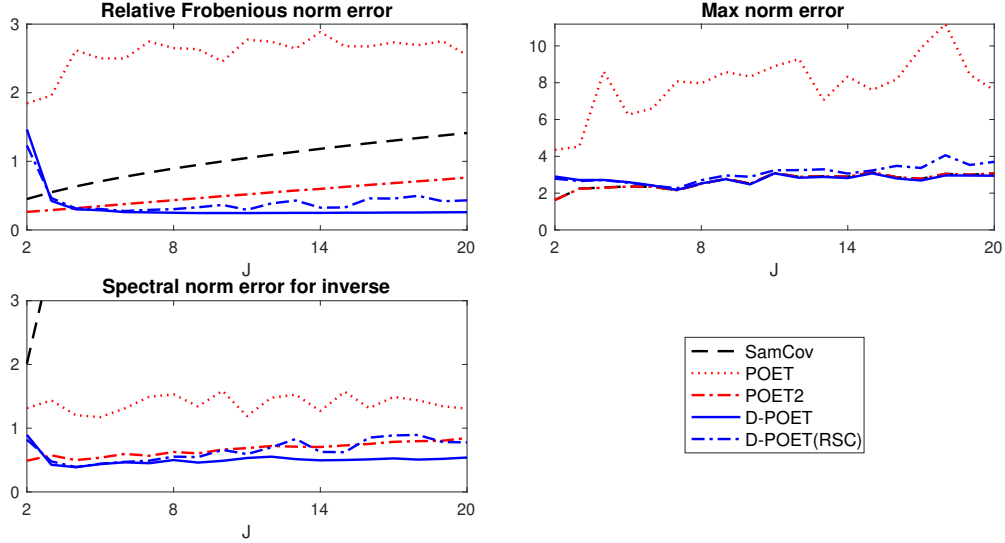


Figure 2: Averages of  $\|\widehat{\Sigma} - \Sigma\|_{\Sigma}$ ,  $\|\widehat{\Sigma} - \Sigma\|_{\max}$  and  $\|(\widehat{\Sigma})^{-1} - \Sigma^{-1}\|$  for Double-POET, Double-POET(RSC), POET, POET2 and the sample covariance matrix against  $J$  with a fixed  $p_j = 30$ .

For comparison, we employed Double-POET, POET, and the sample covariance matrix (SamCov) to estimate the true covariance matrix of  $y$ ,  $\Sigma$ . The estimation errors are measured in the following norms:  $\|\widehat{\Sigma} - \Sigma\|_{\Sigma}$ ,  $\|\widehat{\Sigma} - \Sigma\|_{\max}$ , and  $\|(\widehat{\Sigma})^{-1} - \Sigma^{-1}\|$ , where  $\widehat{\Sigma}$  is one of the covariance matrix estimators. For the POET estimation, we estimated the covariance matrix with two different numbers of factors: (i) POET uses the  $k$  number of factors, and (ii) POET2 uses  $k + r$  factors, where  $r = J \times r_j$ . For Double-POET, we considered two different cases: (i) D-POET with the known group membership, and (ii) D-POET(RSC) suggested in Section 3.5 with the unknown group membership. The proposed numbers of global and local factors estimation method in Section 3.4 is applied with  $k_{\max} = 10 + r$  and  $r_{j,\max} = 10$  for each estimation. In addition, we employed the soft thresholding scheme for both POET and Double-POET.

Figures 1 and 2 plot the averages of  $\|\widehat{\Sigma} - \Sigma\|_{\Sigma}$ ,  $\|\widehat{\Sigma} - \Sigma\|_{\max}$  and  $\|(\widehat{\Sigma})^{-1} - \Sigma^{-1}\|$  from different methods against  $p$  and  $J$ , respectively. From Figure 1, we find that D-POET performs the best. When comparing the POET-type procedures, POET2 performs better than POET. This may be because POET ignores important local factors. D-POET(RSC)

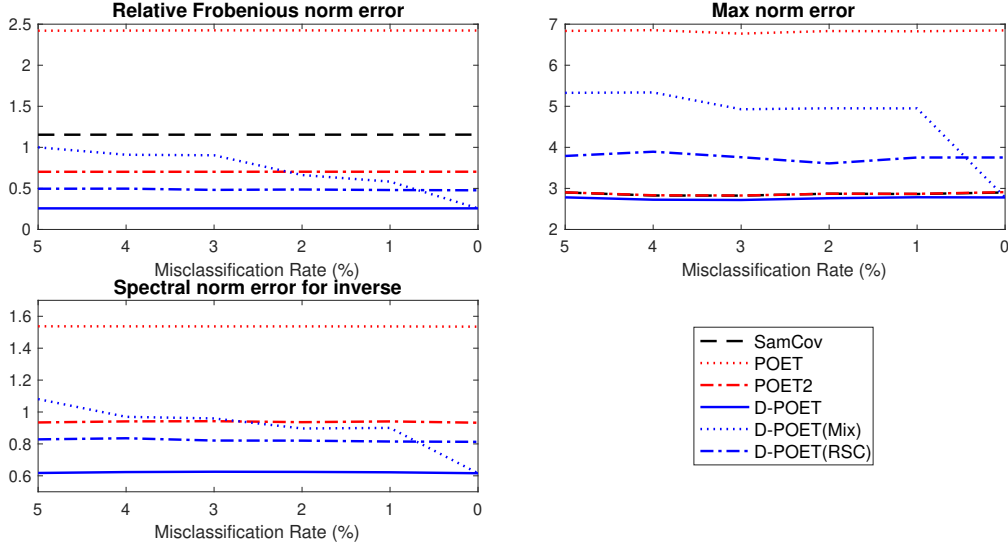


Figure 3: Averages of  $\|\widehat{\Sigma} - \Sigma\|_{\Sigma}$ ,  $\|\widehat{\Sigma} - \Sigma\|_{\max}$  and  $\|(\widehat{\Sigma})^{-1} - \Sigma^{-1}\|$  for Double-POET, Double-POET(RSC), Double-POET(Mix), POET, POET2 and the sample covariance matrix against the misclassification rate with fixed  $J = 20$  and  $p_j = 20$ .

with the unknown group membership performs better than POET2 under different norms. This confirms that the RSC method in Section 3.5 can detect the membership well. Under the max norm, all estimators except POET perform roughly the same. This is because the thresholding or imposing the local factor structure affects mainly the elements of the covariance matrix that are nearly zero, and the elementwise norm depicts the magnitude of the largest elementwise absolute error. Figure 2 shows the similar results except when  $J$  is extremely small. This may be because for the small  $J$  ( $J = 2$ ), we have weak global factors rather than local factors. We note that the estimation errors of POET2 increase as  $J$  grows because the estimator includes more noises on the off-diagonal blocks of the local factor component. The above results support the theoretical findings presented in Section 3.

We further explored the performance of Double-POET when the group membership is misclassified. In Figure 3, we report the average errors under different norms against the misclassified rate of the group membership for Double-POET(Mix) with  $J = 20$  and  $p_j = 20$ . Figure 3 implies that Double-POET method performs better than POET2 unless the misclassification rate is greater than 3%.

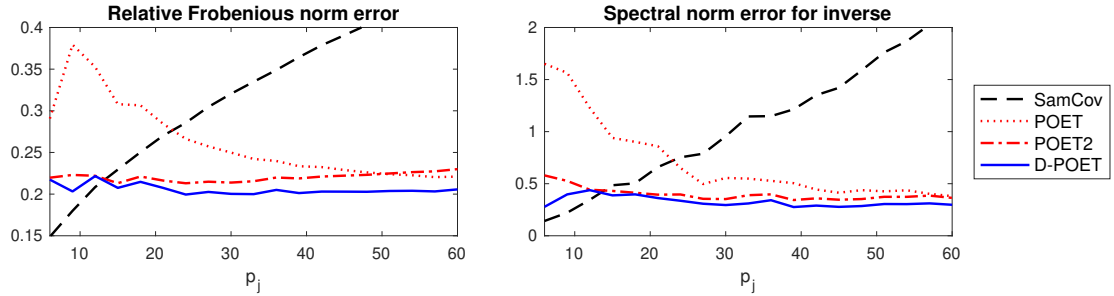


Figure 4: Averages of  $\|\widehat{\Sigma}^j - \Sigma^j\|_{\Sigma^j}$  and  $\|(\widehat{\Sigma}^j)^{-1} - (\Sigma^j)^{-1}\|$  for Double-POET, POET, POET2 and the sample covariance matrix against  $p_j$  with a fixed  $J = 10$ .

We now demonstrate the blessing of dimensionality using Double-POET when estimating the local covariance matrix  $\Sigma^j$  and its inverse. We generated the data as above. The SamCov and POET estimators are obtained using the group sample (i.e.,  $T \times p_j$  observation matrix). For the POET method, we used the number of factors as  $k + r_j = 5$  in each estimation. The Double-POET and POET2 estimators for  $\Sigma^j$  are the  $j$ th diagonal block of  $\widehat{\Sigma}^{\mathcal{D}}$  and  $\widehat{\Sigma}_2^{\mathcal{T}}$ , respectively. We calculated  $\|\widehat{\Sigma}^j - \Sigma^j\|_{\Sigma^j}$  and  $\|(\widehat{\Sigma}^j)^{-1} - (\Sigma^j)^{-1}\|$  for  $j = 1$ . Note that we do not present the results of max norm,  $\|\widehat{\Sigma}^j - \Sigma^j\|_{\max}$ , as all estimators perform very similar to that shown in Figure 1.

Figure 4 depicts the averages of  $\|\widehat{\Sigma}^j - \Sigma^j\|_{\Sigma^j}$  and  $\|(\widehat{\Sigma}^j)^{-1} - (\Sigma^j)^{-1}\|$  for Double-POET, POET, and the sample covariance matrix against  $p_j$  with a fixed  $J = 10$ , while Figure 5 plots their average errors against  $J$  with fixed  $p_j = 30$ . Figures 4 and 5 show that Double-POET has smaller estimation errors than other methods under different norms. In Figure 4, Double-POET significantly outperforms POET when  $p_j$  is small, while the estimation error gap between Double-POET and POET decreases as  $p_j$  grows. The estimation error of POET2 under the relative Frobenius norm tends to increase as  $p_j$  grows. This is because, even though the global factor component can be estimated more accurately, the estimator includes redundant information on the local factor part. In contrast, Figure 5 shows that, except when  $J$  is very small, the Double-POET method constantly dominates the other methods. This might be because by using other local groups' information, Double-POET

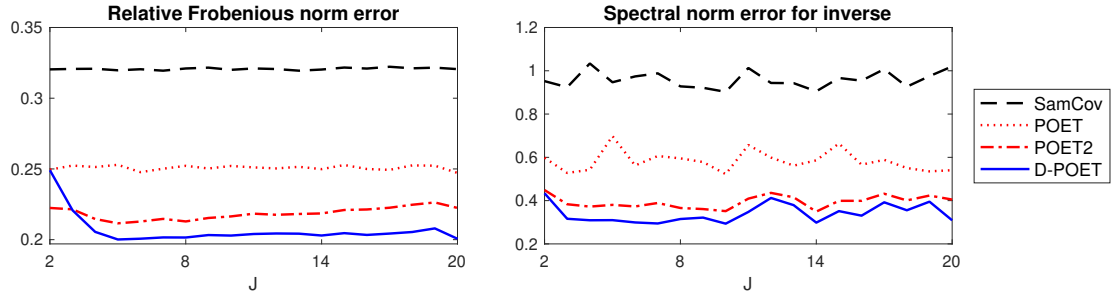


Figure 5: Averages of  $\|\widehat{\Sigma}^j - \Sigma^j\|_{\Sigma^j}$  and  $\|(\widehat{\Sigma}^j)^{-1} - (\Sigma^j)^{-1}\|$  for Double-POET, POET, POET2 and the sample covariance matrix against  $J$  with a fixed  $p_j = 30$ .

can estimate global factors better than POET, especially when  $p_j$  is small. That is, the proposed Double-POET enjoys the blessing of dimensionality, which corresponds to the theoretical analysis discussed in Section 3.3.

We also examined the performance of the modified eigenvalue ratio (MER) estimator introduced in Section 3.4 for the number of global factors. For comparison, we employed alternative estimators: the  $\text{BIC}_3$  estimator of Bai and Ng (2002), the ED estimator of Onatski (2010), the ER estimator of Ahn and Horenstein (2013), and the estimator of Alessi et al. (2010) (ABC). We used the same data generating process as before but considered different number of global factors,  $k \in \{0, 3, 6\}$ , and sample sizes,  $p \in \{100, 200, 300\}$  and  $T \in \{150, 300\}$ . We set  $k_{\max} = 10 + r$  and  $\varphi_p = 0.3 \log p$  for MER estimator, and  $k_{\max} = 10$  for the other estimators. We calculated the average of  $\hat{k}$  and the percentage correctly determining  $k$  over 500 replications.

The results are reported in Table 1. We find that the proposed MER method performs the best across the different number of global factors. In particular, when  $k = 0$ , MER and ED tend to estimate correctly, while other estimators tend to overestimate  $k$ . We note that ER does not include the case of  $k = 0$ . When  $k > 0$ , all estimators except  $\text{BIC}_3$  perform well, but MER and ER slightly outperform ED and ABC. MER often underestimates  $k$  if both  $p$  and  $T$  are small, but it selects  $k$  correctly as  $p$  grows. The results demonstrate that the proposed MER method is appropriate for the model selection problem for both the

Table 1: Average estimates of the number of global factors.

$p$	$T$	MER	BIC <sub>3</sub>	ED	ER	ABC
$k = 0$						
100	150	0.26 (0.91)	3.69 (0.00)	0.31 (0.82)	2.59 (0.00)	0.62 (0.64)
200	150	0.00 (1.00)	3.22 (0.00)	0.05 (0.96)	3.39 (0.00)	0.94 (0.55)
300	150	0.00 (1.00)	4.20 (0.00)	0.09 (0.93)	3.78 (0.00)	0.90 (0.55)
100	300	0.05 (0.98)	2.71 (0.00)	0.13 (0.93)	3.41 (0.00)	0.21 (0.83)
200	300	0.00 (1.00)	5.20 (0.00)	0.02 (0.98)	3.40 (0.00)	0.55 (0.68)
300	300	0.00 (1.00)	7.44 (0.00)	0.22 (0.79)	2.35 (0.00)	0.42 (0.75)
$k = 3$						
100	150	2.09 (0.50)	3.56 (0.47)	3.03 (0.98)	3.00 (1.00)	3.16 (0.86)
200	150	2.99 (1.00)	4.31 (0.02)	3.11 (0.90)	3.00 (1.00)	3.12 (0.90)
300	150	3.00 (1.00)	4.62 (0.03)	3.01 (0.99)	3.00 (1.00)	3.20 (0.84)
100	300	2.92 (0.95)	3.29 (0.71)	3.01 (0.99)	3.00 (1.00)	3.03 (0.98)
200	300	3.00 (1.00)	5.53 (0.00)	3.02 (0.98)	3.00 (1.00)	3.06 (0.95)
300	300	3.00 (1.00)	6.83 (0.00)	3.03 (0.97)	3.00 (1.00)	3.11 (0.91)
$k = 6$						
100	150	5.53 (0.77)	6.03 (0.97)	4.67 (0.76)	5.75 (0.85)	6.14 (0.87)
200	150	5.96 (0.99)	6.63 (0.41)	6.00 (1.00)	6.00 (1.00)	6.08 (0.93)
300	150	6.00 (1.00)	6.21 (0.79)	6.02 (0.98)	6.00 (1.00)	6.03 (0.97)
100	300	6.00 (1.00)	6.00 (1.00)	6.05 (0.98)	6.00 (1.00)	6.01 (0.99)
200	300	6.00 (1.00)	6.46 (0.55)	6.00 (1.00)	6.00 (1.00)	6.02 (0.98)
300	300	6.00 (1.00)	8.04 (0.00)	6.00 (1.00)	6.00 (1.00)	6.01 (0.99)

**Note:** Each entry depicts the average of  $\hat{k}$  over 500 replications with the number in parentheses indicating the percentage correctly estimating  $k$ , the number of global factors. We set  $k_{\max} = 10 + r$  and  $\varphi_p = 0.3 \log p$  for MER estimator, and  $k_{\max} = 10$  for the other estimators. The number of group is fixed at  $J = 10$ .

multi-level factor model and the single-level factor model.

## 5 Application to Portfolio Allocation

In this section, we applied the proposed Double-POET method to a minimum variance portfolio allocation study using global stock data. We collected daily transaction prices of international stock markets over 20 countries by the total market capitalization from January 2, 2016, to December 31, 2021. We selected top 100 firms at most for each country based on the market cap and used weekly log-returns to mitigate the effect of different trading hours.

Table 2: Data Distribution

Country (code)	The number of firms	Country (code)	The number of firms
Australia (AU)	100	Japan (JP)	100
Brazil (BR)	95	Netherlands (NL)	65
Canada (CA)	100	Singapore (SG)	100
China (CN)	77	South Africa (ZA)	100
France (FR)	100	South Korea (KR)	100
Germany (DE)	100	Spain (ES)	61
Hong Kong (HK)	100	Switzerland (CH)	100
India (IN)	100	Thailand (TH)	100
Indonesia (ID)	94	United Kingdom (GB)	100
Italy (IT)	100	United States (US)	100
		Total	1892

We excluded stocks with missing returns and no variation in this period. This operation leads to total 1892 stocks for this period. The distribution of our sample is presented in Table 2.

We calculated the Double-POET, POET, and SamCov estimators for each month. In both Double-POET and POET procedures, we estimated the idiosyncratic volatility matrix based on 11 Global Industrial Classification Standard (GICS) sectors (Ait-Sahalia and Xiu, 2017; Fan et al., 2016). For example, the idiosyncratic components for the different sectors were set to zero, and we maintained these for the same sector. This location-based thresholding preserves positive definiteness and corresponds to the hard-thresholding scheme with sector information. We varied the number of (global) factors  $k$  from 1 to 5 for both Double-POET and POET. To determine the number of local factors for Double-POET, we used the eigenvalue ratio method suggested by Ahn and Horenstein (2013) with  $r_{j,\max} = 10$ . We also considered POET2 using  $k + \sum_{j=1}^{20} \hat{r}_j$  number of factors from the best performing Double-POET estimator.

To analyze the out-of-sample portfolio allocation performance, we considered the following constrained minimum variance portfolio allocation problem (Fan et al., 2012; Jagan-

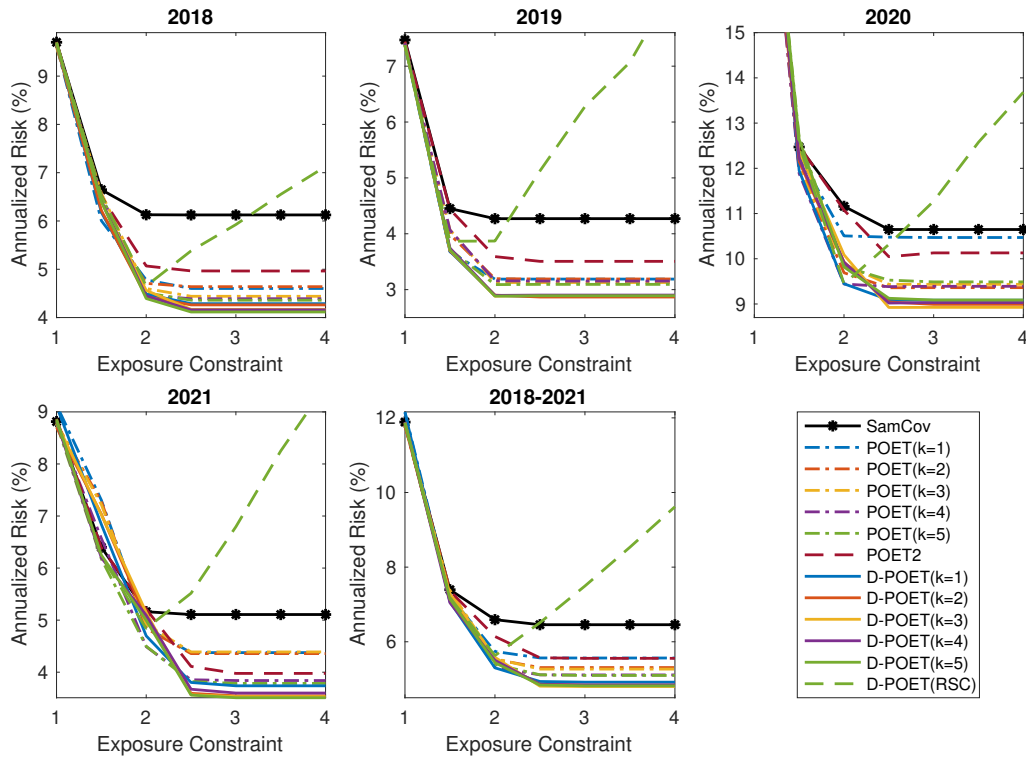


Figure 6: Out-of-sample risks of the optimal portfolios constructed by the SamCov, POET, POET2, Double-POET and Double-POET(RSC) estimators for the global stock market.

nathan and Ma, 2003):

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

where  $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^p$ , the gross exposure constraint  $c$  was varied from 1 to 4, and  $\widehat{\Sigma}$  is one of the volatility matrix estimators from Double-POET, POET, and SamCov. We constructed the portfolio at the beginning of each month, based on the stock weights calculated using the data from the past 24 months ( $T = 104$ ). We then held the portfolio for one month and calculated the square root of the realized volatility using the weekly portfolio log-returns. Their average was used for the out-of-sample risk. We considered five out-of-sample periods: 2018, 2019, 2020, 2021 and the whole period (2018-2021).

Figure 6 depicts the out-of-sample risks of the portfolios constructed by SamCov, POET,

POET2, Double-POET, and Double-POET(RSC) with  $k = 5$  against the exposure constraint. From Figure 6, we find that Double-POET outperforms POET for the same number of factors  $k$ , while  $k = 3$  and  $k = 5$  yields the best performances for Double-POET and POET, respectively. Double-POET( $k = 3$ ) reduces the minimum risks by 4.3%–6.7% compared to POET( $k = 5$ ). We confirmed that for the purpose of portfolio allocation based on the global stock market, the Double-POET based on the global and national factor model outperforms POET based on the single-level factor model. Thus, we can conjecture that incorporating the latent global and national factor models helps account for the global market dynamics. We note that Double-POET(RSC) does not perform well due to misclassified group membership. One possible explanation is that there might be other type of local factors in addition to the nation-specific factors. For example, if the industrial risk and the national risk are nested on the idiosyncratic part after removing the global factors, the suggested RSC method cannot properly detect only the national group membership. This is an interesting research topic to control the unknown nested local groups, so we leave it for a future study.

We also conducted the country-wise volatility matrix estimation based on SamCov, POET, POET2, and Double-POET and applied them to the same portfolio allocation problem as before. Specifically, we used the sample of each country for POET and SamCov procedures. The Double-POET and POET2 estimators for each local covariance matrix are obtained by extracting each diagonal block of the large Double-POET and POET2 estimators, respectively. Figure 7 presents the portfolio behavior of the top 100 stocks in the US. Double-POET shows stable results and reduces the minimum risks by 1.8%–4.0% compared to POET except the period 2021. We note that a powerful bull market lasted in 2021, and the risks could be sufficiently explained by only the first principal component (i.e., the market factor), so that, POET( $k = 1$ ) performs the best in this period. Nevertheless, the overall results indicate that the proposed Double-POET method enjoys the blessing of dimensionality.

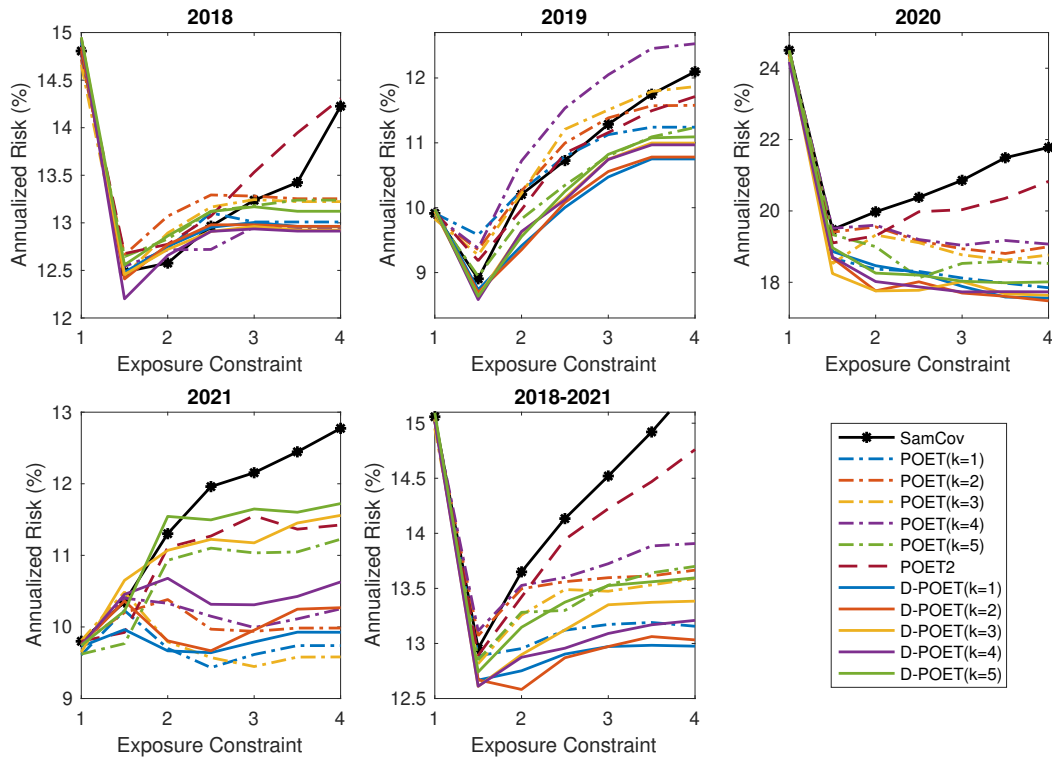


Figure 7: Out-of-sample risks of the optimal portfolios constructed by the SamCov, POET, POET2 and Double-POET estimators for the US stock market.

Figure 8 shows the results of other 19 countries for the whole period. Except for five countries (GB, IN, IT, TH, and ZA), Double-POET outperforms POET. We note that the SamCov estimator sometimes outperforms others for a few countries. Overall, these results indicate that Double-POET can accurately estimate the global factors by harnessing other countries' observations.

## 6 Conclusion

This paper proposes a novel large volatility matrix inference procedure based on the latent global and national factor models. We show the asymptotic behaviors of the proposed Double-POET method and discuss its blessing of dimensionality and efficiency of estimating a large volatility matrix compared to the regular POET procedure. To determine the number

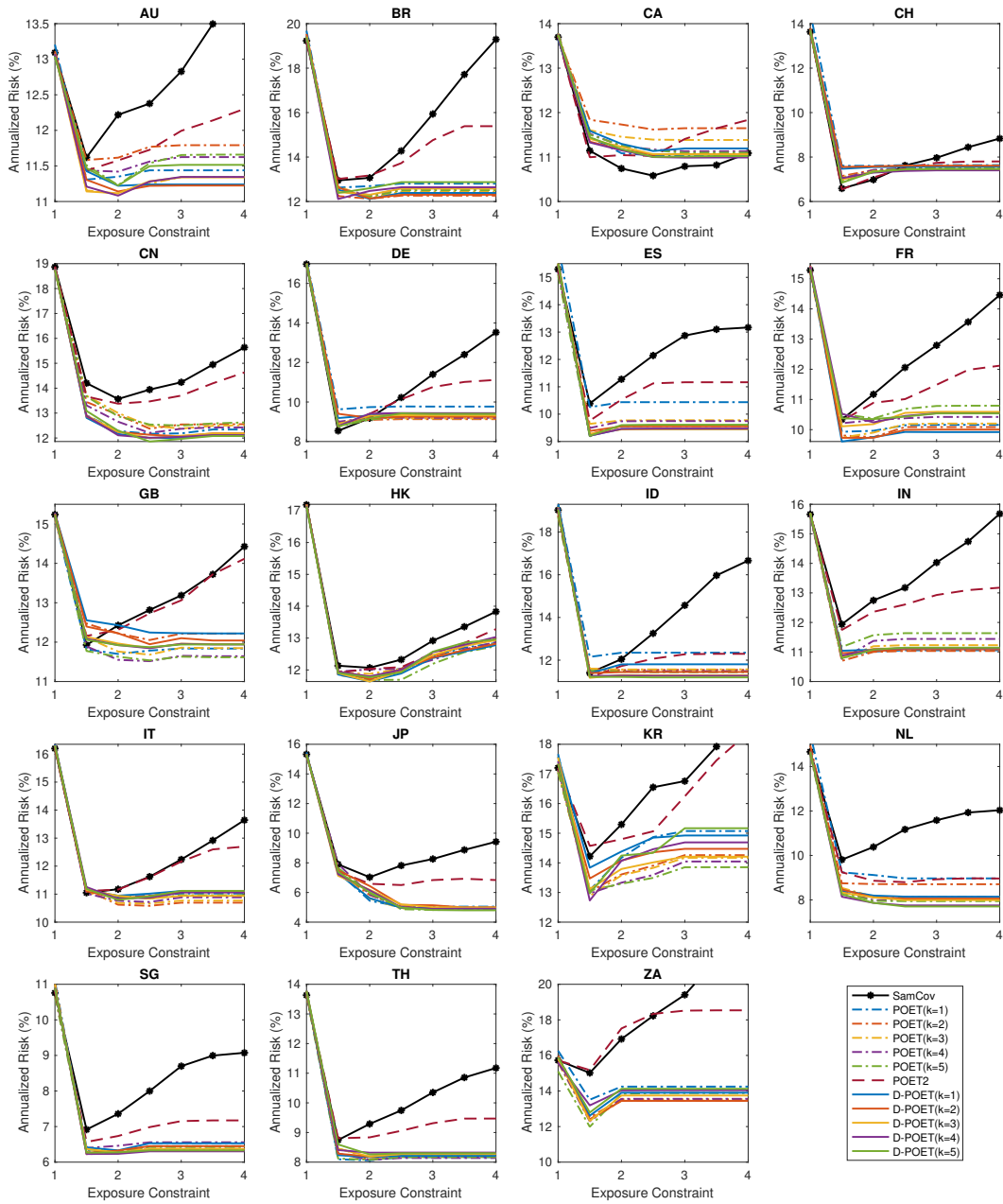


Figure 8: Out-of-sample risks of the optimal portfolios constructed by the SamCov, POET, POET2 and Double-POET estimators for each country's stock market. The out-of-sample period is from 2018 to 2021.

of global factors, we extend the eigenvalue ratio procedure (Ahn and Horenstein, 2013). In addition, when the membership of the local factors is unknown, we suggest the regularized spectral clustering method to find the latent local structure.

In the empirical study, in terms of portfolio allocation, the proposed estimator shows the best performance. It confirms the presence of the national factor structure in global financial markets, which provides the theoretical basis for employing the Double-POET method. In addition, for the country-wise covariance matrix estimation, the Double-POET procedure yields the blessing of dimensionality by accurately estimating the latent global factors using the information outside the local group.

In this paper, we focus on the national risk factor as only the local factor. However, in practice, there could be other types of risk factors that are nested in the local level. Thus, it is interesting and important to develop a large volatility matrix estimation procedure based on unknown-membership local factor models with the nested local-level group factors. We leave this for a future study.

## Acknowledgments

The authors thank the Editor Professor Torben Andersen, the Associate Editor, and two referees for their careful reading of this paper and valuable comments. The research of Donggyu Kim was supported in part by the National Research Foundation of Korea (NRF) (2021R1C1C1003216).

## 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.
- ALESSI, L., M. BARIGOZZI, AND M. CAPASSO (2010): “Improved penalization for determining the number of factors in approximate factor models,” *Statistics & Probability Letters*, 80, 1806–1813.
- AMINI, A. A., A. CHEN, P. J. BICKEL, AND E. LEVINA (2013): “Pseudo-likelihood methods for community detection in large sparse networks,” *The Annals of Statistics*, 41, 2097–2122.
- 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.
- ARUOBA, S. B., F. X. DIEBOLD, M. A. KOSE, AND M. E. TERRONES (2011): “Globalization, the business cycle, and macroeconomic monitoring,” in *NBER international seminar on macroeconomics*, University of Chicago Press Chicago, IL, vol. 7, 245–286.
- 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.

- 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 A. CHEN (2009): “A nonparametric view of network models and Newman–Girvan and other modularities,” *Proceedings of the National Academy of Sciences*, 106, 21068–21073.
- BICKEL, P. J. AND E. LEVINA (2008): “Covariance regularization by thresholding,” *The Annals of Statistics*, 36, 2577–2604.
- BOIVIN, J. AND S. NG (2006): “Are more data always better for factor analysis?” *Journal of Econometrics*, 132, 169–194.
- CAI, T. AND W. LIU (2011): “Adaptive thresholding for sparse covariance matrix estimation,” *Journal of the American Statistical Association*, 106, 672–684.
- CAI, T. T. AND X. LI (2015): “Robust and computationally feasible community detection in the presence of arbitrary outlier nodes,” *The Annals of Statistics*, 43, 1027–1059.
- CAI, T. T., Z. MA, AND Y. WU (2013): “Sparse PCA: Optimal rates and adaptive estimation,” *The Annals of Statistics*, 41, 3074–3110.
- CANDÈS, E. J. AND B. RECHT (2009): “Exact matrix completion via convex optimization,” *Foundations of Computational mathematics*, 9, 717–772.
- CHAMBERLAIN, G. AND M. ROTHSCHILD (1983): “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets,” *Econometrica*, 51, 1281–1304.
- CHAUDHURI, K., F. C. GRAHAM, AND A. TSIATAS (2012): “Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model.” in *COLT*, vol. 23, 35–1.

- 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.
- FAN, J., Y. FAN, AND J. LV (2008): “High dimensional covariance matrix estimation using a factor model,” *Journal of Econometrics*, 147, 186–197.
- 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 (2018): “Robust high-dimensional volatility matrix estimation for high-frequency factor model,” *Journal of the American Statistical Association*, 113, 1268–1283.
- (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,” *The 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,” *The Annals of Statistics*, 46, 1383.
- FAN, J., W. WANG, AND Y. ZHONG (2018b): “An  $\infty$  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,” *The 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.
- GIGLIO, S. AND D. XIU (2021): “Asset pricing with omitted factors,” *Journal of Political Economy*, 129, 1947–1990.
- GIRVAN, M. AND M. E. NEWMAN (2002): “Community structure in social and biological networks,” *Proceedings of the national academy of sciences*, 99, 7821–7826.
- GREGORY, A. W. AND A. C. HEAD (1999): “Common and country-specific fluctuations in productivity, investment, and the current account,” *Journal of Monetary Economics*, 44, 423–451.
- HAGEN, L. AND A. B. KAHNG (1992): “New spectral methods for ratio cut partitioning and clustering,” *IEEE transactions on computer-aided design of integrated circuits and systems*, 11, 1074–1085.
- HAJEK, B., Y. WU, AND J. XU (2015): “Achieving exact cluster recovery threshold via semidefinite programming: Extensions,” *arXiv preprint arXiv:1502.07738*.
- HALLIN, M. AND R. LIŠKA (2011): “Dynamic factors in the presence of blocks,” *Journal of Econometrics*, 163, 29–41.
- HAN, X. (2021): “Shrinkage estimation of factor models with global and group-specific factors,” *Journal of Business & Economic Statistics*, 39, 1–17.
- JAGANNATHAN, R. AND T. MA (2003): “Risk reduction in large portfolios: Why imposing the wrong constraints helps,” *The Journal of Finance*, 58, 1651–1683.

- JOHNSTONE, I. M. AND A. Y. LU (2009): “On consistency and sparsity for principal components analysis in high dimensions,” *Journal of the American Statistical Association*, 104, 682–693.
- JOSEPH, A., B. YU, ET AL. (2016): “Impact of regularization on spectral clustering,” *The Annals of Statistics*, 44, 1765–1791.
- 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.
- KIM, D. AND J. FAN (2019): “Factor GARCH-Itô models for high-frequency data with application to large volatility matrix prediction,” *Journal of Econometrics*, 208, 395–417.
- 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.
- LEI, J. AND A. RINALDO (2013): “Consistency of spectral clustering in sparse stochastic block models,” *arXiv preprint arxiv:1312.2050*.
- MA, Z. (2013): “Sparse principal component analysis and iterative thresholding,” *The Annals of Statistics*, 41, 772–801.
- MCSHERRY, F. (2001): “Spectral partitioning of random graphs,” in *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, IEEE, 529–537.
- MOENCH, E., S. NG, AND S. POTTER (2013): “Dynamic hierarchical factor models,” *Review of Economics and Statistics*, 95, 1811–1817.
- NEGAHBAN, S. AND M. J. WAINWRIGHT (2011): “Estimation of (near) low-rank matrices with noise and high-dimensional scaling,” *The Annals of Statistics*, 39, 1069–1097.

- ONATSKI, A. (2010): “Determining the number of factors from empirical distribution of eigenvalues,” *The Review of Economics and Statistics*, 92, 1004–1016.
- QIN, T. AND K. ROHE (2013): “Regularized spectral clustering under the degree-corrected stochastic blockmodel,” in *Advances in Neural Information Processing Systems*, 3120–3128.
- ROHE, K., S. CHATTERJEE, AND B. YU (2011): “Spectral clustering and the high-dimensional stochastic blockmodel,” *The Annals of Statistics*, 1878–1915.
- ROTHMAN, A. J., E. LEVINA, AND J. ZHU (2009): “Generalized thresholding of large covariance matrices,” *Journal of the American Statistical Association*, 104, 177–186.
- SHI, J. AND J. MALIK (2000): “Normalized cuts and image segmentation,” *IEEE Transactions on pattern analysis and machine intelligence*, 22, 888–905.
- SHIN, M., D. KIM, AND J. FAN (2021): “Adaptive robust large volatility matrix estimation based on high-frequency financial data,” *arXiv preprint arXiv:2102.12752*.
- 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.
- TRAPANI, L. (2018): “A randomized sequential procedure to determine the number of factors,” *Journal of the American Statistical Association*, 113, 1341–1349.
- 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,” *The Annals of Statistics*, 45, 1342.

# A Appendix

## A.1 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  $\Sigma$  in decreasing order. Let  $\{\bar{\delta}_i, \bar{v}_i\}_{i=1}^k$  be the leading eigenvalues and eigenvectors of  $\mathbf{B}\mathbf{B}'$  and the rest zero. Similarly, for each group  $j$ , let  $\{\kappa_i^j, \eta_i^j\}_{i=1}^{p_j}$  be the eigenvalues and eigenvectors of  $\Sigma_E^j$  in decreasing order, and  $\{\bar{\kappa}_i^j, \bar{\eta}_i^j\}_{i=1}^{r_j}$  for  $\Lambda^j \Lambda^{j'}$ .

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

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

$$|\delta_i - \bar{\delta}_i| \leq \|\Sigma_E\| \text{ for } i \leq k, \quad |\delta_i| \leq \|\Sigma_E\| \text{ for } i > k,$$

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

$$|\kappa_i^j - \bar{\kappa}_i^j| \leq \|\Sigma_u^j\| \text{ for } i \leq r_j, \quad |\kappa_i^j| \leq \|\Sigma_u^j\| \text{ for } i > r_j,$$

and, for  $i \leq r_j$ ,  $\bar{\kappa}_i^j/p_j^{a_2}$  is strictly bigger than zero for all  $p_j$ .

The following lemma presents the individual convergence rate of leading eigenvectors using Lemma A.1 and the  $l_\infty$  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*

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

(ii) *For each group  $j$ , we have*

$$\max_{i \leq r_j} \|\bar{\eta}_i^j - \eta_i^j\|_\infty \leq C \frac{\|\Sigma_u^j\|_\infty}{p_j^{3(a_2 - \frac{1}{2})}}.$$

*Proof.* (i) Let  $\mathbf{B} = (\tilde{b}_1, \dots, \tilde{b}_k)$ . 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  $\tilde{\mathbf{V}} = (\bar{v}_1, \dots, \bar{v}_k)$ , the coherence  $\mu(\tilde{\mathbf{V}}) = p \max_i \sum_{j=1}^k \tilde{\mathbf{V}}_{ij}^2/k \leq Cp^{1-a_1}$ , where  $\tilde{\mathbf{V}}_{ij}$  is the  $(i, j)$  entry of  $\tilde{\mathbf{V}}$ . Thus, by Theorem 1 of Fan et al. (2018b), we have

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

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

**Lemma A.3.** *Under Assumption 3.1, for  $i \leq k$ , we have*

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

*Proof.* The first statement is followed by Assumption 3.1 and Weyl's theorem. We have

$$\hat{\Sigma} = \mathbf{B}\mathbf{B}' + \Lambda\Lambda' + \Sigma_u + (\hat{\Sigma} - \Sigma) = \mathbf{B}\mathbf{B}' + \Sigma_E + (\hat{\Sigma} - \Sigma).$$

We can treat  $\mathbf{B}\mathbf{B}'$  as a low rank matrix and the remaining terms as a perturbation matrix.

By Theorem 1 of Fan et al. (2018b), Lemma A.2 and Assumption 3.1, we have

$$\begin{aligned} \|\hat{v}_i - v_i\|_\infty &\leq Cp^{2(1-a_1)} \frac{\|\Sigma_E + (\hat{\Sigma} - \Sigma)\|_\infty}{p^{a_1}\sqrt{p}} \leq Cp^{2(1-a_1)} \frac{\|\Sigma_E\|_\infty}{p^{a_1}\sqrt{p}} + Cp^{2(1-a_1)} \frac{\|\hat{\Sigma} - \Sigma\|_{\max}}{p^{a_1-1}\sqrt{p}} \\ &= O_P\left(\frac{\|\Sigma_E\|_\infty}{p^{3(a_1-\frac{1}{2})}} + p^{3(1-a_1)} \sqrt{\frac{\log p}{Tp}}\right). \end{aligned}$$

$\square$

**Lemma A.4.** *Under the assumptions of Theorem 3.1, for  $i \leq r_j$ , we have*

$$|\widehat{\kappa}_i^j / \kappa_i^j - 1| = O_P \left( p^{\frac{5}{2}(1-a_1)+c(1-a_2)} \sqrt{\log p/T} + 1/p^{\frac{5a_1}{2}-\frac{3}{2}-2c+ca_2} \right),$$

$$\|\widehat{\eta}_i^j - \eta_i^j\|_\infty = O_P \left( p^{\frac{5}{2}(1-a_1)+c(\frac{5}{2}-3a_2)} \sqrt{\frac{\log p}{T}} + \frac{1}{p^{\frac{5a_1}{2}-\frac{3}{2}+c(3a_2-\frac{7}{2})}} + \frac{m_p}{p^{3c(a_2-\frac{1}{2})}} \right).$$

*Proof.* We have

$$\|\Sigma_E\| \leq \|\Lambda\Lambda'\| + \|\Sigma_u\| \leq \|\Lambda\Lambda'\| + O(m_p) = O(p^{ca_2}).$$

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

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

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.3, we have

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

Thus, we have

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

Then, we have

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

Therefore, the first statement is followed by (A.2) and the Weyl's theorem.

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

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

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

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

□

**Proof of Theorem 3.1.** We first consider (3.1). By definition,  $\|\widehat{\mathbf{\Sigma}}_u^{\mathcal{D}} - \widehat{\mathbf{\Sigma}}_u\|_{\max} = \max_{ij} |s_{ij}(\widehat{\sigma}_{ij}) - \widehat{\sigma}_{ij}| \leq \max_{ij} \tau_{ij} = O_P(\tau)$ . Hence, it suffices to show  $\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\|_{\max} = O_P(\omega_T)$ .

We have

$$\|\widehat{\mathbf{\Sigma}}_u - \mathbf{\Sigma}_u\|_{\max} \leq \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} + \|\widehat{\mathbf{V}}\widehat{\mathbf{\Gamma}}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max} + \|\widehat{\mathbf{\Phi}}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Phi}}' - \mathbf{\Lambda}\mathbf{\Lambda}'\|_{\max}.$$

By Assumption 3.1(iii) and (A.1), we have

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

For  $\|\widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' - \Lambda\Lambda'\|_{\max}$ , we have

$$\|\widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' - \Lambda\Lambda'\|_{\max} = \max_j \|\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Lambda^j\Lambda^{j'}\|_{\max}.$$

For each group  $j$ , let  $\Lambda^j\Lambda^{j'} = \widetilde{\Phi}^j\widetilde{\Psi}^j\widetilde{\Phi}^{j'}$ , where  $\widetilde{\Psi}^j = \text{diag}(\bar{\kappa}_1^j, \dots, \bar{\kappa}_{r_j}^j)$  and the corresponding eigenvectors  $\widetilde{\Phi}^j = (\bar{\eta}_1, \dots, \bar{\eta}_{r_j})$ . In addition, let  $\Psi^j = \text{diag}(\kappa_1^j, \dots, \kappa_{r_j}^j)$  and  $\Phi^j = (\eta_1, \dots, \eta_{r_j})$  to be the leading eigenvalues and the corresponding eigenvectors of  $\Sigma_E^j$ , respectively. Then, we have

$$\begin{aligned} \|\Phi^j\Psi^{j\frac{1}{2}} - \widetilde{\Phi}^j\widetilde{\Psi}^{j\frac{1}{2}}\|_{\max} &\leq \|\Lambda^j\widetilde{\Psi}^{j-\frac{1}{2}}(\Psi^{j\frac{1}{2}} - \widetilde{\Psi}^{j\frac{1}{2}})\|_{\max} + \|(\Phi^j - \widetilde{\Phi}^j)\Psi^{j\frac{1}{2}}\|_{\max} \\ &\leq \frac{\|\Sigma_u^j\|}{p_j^{a_2}} + \frac{\|\Sigma_u^j\|_{\infty}}{\sqrt{p_j^{5a_2-3}}} = o(1). \end{aligned} \quad (\text{A.4})$$

Since  $\|\Lambda^j\|_{\max} = \|\widetilde{\Phi}^j\widetilde{\Psi}^{j\frac{1}{2}}\|_{\max} = O(1)$ ,  $\|\Phi^j\Psi^{j\frac{1}{2}}\|_{\max} = O(1)$  and  $\|\Phi^j\|_{\max} = O(1/\sqrt{p_j^{a_2}})$ .

Using this fact and results from Lemmas A.1, A.2 and A.4, we can show

$$\begin{aligned} \|\widetilde{\Phi}^j\widetilde{\Psi}^j\widetilde{\Phi}^{j'} - \Phi^j\Psi^j\Phi^{j'}\|_{\max} &\leq O(p_j^{-a_2}\|\widetilde{\Psi}^j - \Psi^j\|_{\max} + \sqrt{p_j^{a_2}}\|\widetilde{\Phi}^j - \Phi^j\|_{\max}) = O(m_p/\sqrt{p^{c(5a_2-3)}}), \\ \|\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Phi^j\Psi^j\Phi^{j'}\|_{\max} &\leq O_P(p_j^{-a_2}\|\widehat{\Psi}^j - \Psi^j\|_{\max} + \sqrt{p_j^{a_2}}\|\widehat{\Phi}^j - \Phi^j\|_{\max}) \\ &= O_P(p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1-\frac{3}{2}+c(\frac{5}{2}a_2-\frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}). \end{aligned}$$

By using these rates, we obtain

$$\|\widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' - \Lambda\Lambda'\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1-\frac{3}{2}+c(\frac{5}{2}a_2-\frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}). \quad (\text{A.5})$$

By (A.3) and (A.5), we have

$$\|\widehat{\Sigma}_u - \Sigma_u\|_{\max} = O_P \left( p^{\frac{5}{2}(1-a_1) + \frac{5}{2}c(1-a_2)} \sqrt{\frac{\log p}{T}} + \frac{1}{p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})}} + \frac{m_p}{\sqrt{p^{c(5a_2-3)}}} \right).$$

Therefore,  $\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_{\max} = O_P(\tau + \omega_T) = O_P(\omega_T)$ , when  $\tau$  is chosen as the same order of  $\omega_T = p^{\frac{5}{2}(1-a_1) + \frac{5}{2}c(1-a_2)} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} + c(\frac{5}{2}a_2 - \frac{7}{2})} + m_p/\sqrt{p^{c(5a_2-3)}}$ .

Consider (3.2). Similar to the proofs of Theorem 2.1 in Fan et al. (2011), we can show  $\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_2 = O_P(m_p \omega_T^{1-q})$ . In addition, since  $\lambda_{\min}(\Sigma_u) > c_1$  and  $m_p \omega_T^{1-q} = o(1)$ , the minimum eigenvalue of  $\widehat{\Sigma}_u^{\mathcal{D}}$  is strictly bigger than 0 with probability approaching 1. Then, we have  $\|(\widehat{\Sigma}_u^{\mathcal{D}})^{-1} - \Sigma_u^{-1}\|_2 \leq \lambda_{\min}(\Sigma_u)^{-1} \|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_2 \lambda_{\min}(\widehat{\Sigma}_u^{\mathcal{D}})^{-1} = O_P(m_p \omega_T^{1-q})$ .

Consider (3.3). By the results of (A.1), (A.5) and (3.1), we have

$$\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\max} \leq \|\widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max} + \|\widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' - \Lambda\Lambda'\|_{\max} + \|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_{\max} = O_P(\omega_T).$$

Consider (3.4). Define  $\widehat{\Sigma}_E^{\mathcal{D}} = \widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' + \widehat{\Sigma}_u^{\mathcal{D}}$ . We first show that  $\|(\widehat{\Sigma}_E^{\mathcal{D}})^{-1} - \Sigma_E^{-1}\| = O_P(m_p \omega_T^{1-q})$ . Let  $\widehat{\mathbf{J}} = \widehat{\Psi}^{\frac{1}{2}}\widehat{\Phi}'(\widehat{\Sigma}_u^{\mathcal{D}})^{-1}\widehat{\Phi}\widehat{\Psi}^{\frac{1}{2}}$  and  $\widetilde{\mathbf{J}} = \widetilde{\Psi}^{\frac{1}{2}}\widetilde{\Phi}'\Sigma_u^{-1}\widetilde{\Phi}\widetilde{\Psi}^{\frac{1}{2}}$ . Using the Sherman-Morrison-Woodbury formula, we have

$$\|(\widehat{\Sigma}_E^{\mathcal{D}})^{-1} - \Sigma_E^{-1}\| \leq \|(\widehat{\Sigma}_u^{\mathcal{D}})^{-1} - \Sigma_u^{-1}\| + \Delta_1,$$

where  $\Delta_1 = \|(\widehat{\Sigma}_u^{\mathcal{D}})^{-1}\widehat{\Phi}\widehat{\Psi}^{\frac{1}{2}}(\mathbf{I}_r + \widehat{\mathbf{J}})^{-1}\widetilde{\Psi}^{\frac{1}{2}}\widetilde{\Phi}'(\widehat{\Sigma}_u^{\mathcal{D}})^{-1} - \Sigma_u^{-1}\widetilde{\Phi}\widetilde{\Psi}^{\frac{1}{2}}(\mathbf{I}_r + \widetilde{\mathbf{J}})^{-1}\widetilde{\Psi}^{\frac{1}{2}}\widetilde{\Phi}'\Sigma_u^{-1}\|$ . Then, the right hand side can be bounded by following terms:

$$\begin{aligned} L_1 &= \|((\widehat{\Sigma}_u^{\mathcal{D}})^{-1} - \Sigma_u^{-1})\widetilde{\Phi}\widetilde{\Psi}^{\frac{1}{2}}(\mathbf{I}_r + \widetilde{\mathbf{J}})^{-1}\widetilde{\Psi}^{\frac{1}{2}}\widetilde{\Phi}'\Sigma_u^{-1}\|, \\ L_2 &= \|\Sigma_u^{-1}(\widehat{\Phi}\widehat{\Psi}^{\frac{1}{2}} - \widetilde{\Phi}\widetilde{\Psi}^{\frac{1}{2}})(\mathbf{I}_r + \widetilde{\mathbf{J}})^{-1}\widetilde{\Psi}^{\frac{1}{2}}\widetilde{\Phi}'\Sigma_u^{-1}\|, \\ L_3 &= \|\Sigma_u^{-1}\widetilde{\Phi}\widetilde{\Psi}^{\frac{1}{2}}((\mathbf{I}_r + \widehat{\mathbf{J}})^{-1} - (\mathbf{I}_r + \widetilde{\mathbf{J}})^{-1})\widetilde{\Psi}^{\frac{1}{2}}\widetilde{\Phi}'\Sigma_u^{-1}\|. \end{aligned}$$

By Lemma A.4,  $\|\Phi^j \Psi^{j\frac{1}{2}} - \widehat{\Phi}^j \widehat{\Psi}^{j\frac{1}{2}}\|_{\max} \leq \|\Lambda^j \widehat{\Psi}^{j-\frac{1}{2}}(\Psi^{j\frac{1}{2}} - \widehat{\Psi}^{j\frac{1}{2}})\|_{\max} + \|(\Phi^j - \widehat{\Phi}^j)\Psi^{j\frac{1}{2}}\|_{\max} =$

$O_P(\omega_T)$ , and by (A.4) and (3.2), we then have

$$\begin{aligned}\|\tilde{\Phi}\tilde{\Psi}^{\frac{1}{2}}\| &\leq \max_j \|\tilde{\Phi}^j\tilde{\Psi}^{j\frac{1}{2}}\| = O_P(\sqrt{p^c}), \\ \|\hat{\Phi}\hat{\Psi}^{\frac{1}{2}} - \tilde{\Phi}\tilde{\Psi}^{\frac{1}{2}}\| &\leq \max_j \sqrt{p^c} \|\hat{\Phi}^j\hat{\Psi}^{j\frac{1}{2}} - \tilde{\Phi}^j\tilde{\Psi}^{j\frac{1}{2}}\|_{\max} = O_P(\sqrt{p^c}\omega_T),\end{aligned}$$

and

$$\begin{aligned}\|\hat{\mathbf{J}} - \tilde{\mathbf{J}}\| &\leq \|(\hat{\Psi}^{\frac{1}{2}}\hat{\Phi}' - \tilde{\Psi}^{\frac{1}{2}}\tilde{\Phi}')(\hat{\Sigma}_u^{\mathcal{D}})^{-1}(\hat{\Phi}\hat{\Psi}^{\frac{1}{2}} - \tilde{\Phi}\tilde{\Psi}^{\frac{1}{2}})\| \\ &\quad + \|(\hat{\Psi}^{\frac{1}{2}}\hat{\Phi}' - \tilde{\Psi}^{\frac{1}{2}}\tilde{\Phi}')(\hat{\Sigma}_u^{\mathcal{D}})^{-1}\tilde{\Phi}\tilde{\Psi}^{\frac{1}{2}}\| + \|\tilde{\Psi}^{\frac{1}{2}}\tilde{\Phi}'((\hat{\Sigma}_u^{\mathcal{D}})^{-1} - \Sigma_u^{-1})\tilde{\Phi}\tilde{\Psi}^{\frac{1}{2}}\| \\ &= O_P(p^c m_p \omega_T^{1-q}).\end{aligned}$$

Since  $\lambda_{\min}(\mathbf{I}_r + \tilde{\mathbf{J}}) \geq \lambda_{\min}(\tilde{\mathbf{J}}) \geq \lambda_{\min}(\Sigma_u^{-1})\lambda_{\min}^2(\tilde{\Phi}\tilde{\Psi}^{\frac{1}{2}}) \geq Cp^c$ , we have  $\|(\mathbf{I}_r + \tilde{\mathbf{J}})^{-1}\| = O_P(1/p^c)$ . Then,  $L_1 = O_P(m_p \omega_T^{1-q})$  by (3.2). In addition,  $L_2 = O_P(p^{-c/2}\|\hat{\Phi}\hat{\Psi}^{\frac{1}{2}} - \tilde{\Phi}\tilde{\Psi}^{\frac{1}{2}}\|) = O_P(\omega_T)$  and  $L_3 = O_P(p^c\|(\mathbf{I}_r + \hat{\mathbf{J}})^{-1} - (\mathbf{I}_r + \tilde{\mathbf{J}})^{-1}\|) = O_P(p^{-c}\|\hat{\mathbf{J}} - \tilde{\mathbf{J}}\|) = O_P(m_p \omega_T^{1-q})$ . Thus, we have

$$\Delta_1 = O_P(m_p \omega_T^{1-q}), \tag{A.6}$$

which yields  $\|(\hat{\Sigma}_E^{\mathcal{D}})^{-1} - \Sigma_E^{-1}\| = O_P(m_p \omega_T^{1-q})$ .

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

$$\|(\hat{\Sigma}^{\mathcal{D}})^{-1} - \Sigma^{-1}\| \leq \|(\hat{\Sigma}_E^{\mathcal{D}})^{-1} - \Sigma_E^{-1}\| + \Delta_2,$$

where  $\Delta_2 = \|(\hat{\Sigma}_E^{\mathcal{D}})^{-1}\hat{\mathbf{V}}\hat{\Gamma}^{\frac{1}{2}}(\mathbf{I}_k + \hat{\mathbf{H}})^{-1}\hat{\Gamma}^{\frac{1}{2}}\hat{\mathbf{V}}'(\hat{\Sigma}_E^{\mathcal{D}})^{-1} - \Sigma_E^{-1}\tilde{\mathbf{V}}\tilde{\Gamma}^{\frac{1}{2}}(\mathbf{I}_k + \tilde{\mathbf{H}})^{-1}\tilde{\Gamma}^{\frac{1}{2}}\tilde{\mathbf{V}}'\Sigma_E^{-1}\|$ . By Weyl's inequality, we have  $\lambda_{\min}(\Sigma_E) > c$  since  $\lambda_{\min}(\Sigma_u) > c$  and  $\lambda_{\min}(\mathbf{A}\mathbf{A}') = 0$ . Hence,  $\|\Sigma_E^{-1}\| = O_P(1)$ . By Lemmas A.1-A.3, we have  $\|\hat{\mathbf{V}}\hat{\Gamma}^{\frac{1}{2}} - \tilde{\mathbf{V}}\tilde{\Gamma}^{\frac{1}{2}}\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)}\sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} - c})$ . Similar to the proof of (A.6), we can show  $\Delta_2 = O_P(m_p \omega_T^{1-q})$ . Therefore, we have  $\|(\hat{\Sigma}^{\mathcal{D}})^{-1} - \Sigma^{-1}\| = O_P(m_p \omega_T^{1-q})$ .

Consider (3.5). We derive the rate of convergence for  $\|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma}$ . The SVD decomposition of  $\Sigma$  is

$$\Sigma = (\mathbf{V}_{p \times k} \quad \Phi_{p \times r} \quad \Omega_{p \times (p-k-r)}) \begin{pmatrix} \Gamma_{k \times k} & & \\ & \Psi_{r \times r} & \\ & & \Theta_{(p-k-r) \times (p-k-r)} \end{pmatrix} \begin{pmatrix} \mathbf{V}' \\ \Phi' \\ \Omega' \end{pmatrix}.$$

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

$$\begin{aligned} \|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\Sigma} &\leq p^{-1/2} \left( \|\Sigma^{-1/2}(\widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\Sigma^{-1/2}\|_F \right. \\ &\quad \left. + \|\Sigma^{-1/2}(\widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' - \Lambda\Lambda')\Sigma^{-1/2}\|_F + \|\Sigma^{-1/2}(\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u)\Sigma^{-1/2}\|_F \right) \\ &=: \Delta_G + \Delta_L + \Delta_S \end{aligned}$$

and

$$\Delta_S = O_P(p^{-1/2}\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_F) = O_P(\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_2) = O_P(m_p\omega_T^{1-q}).$$

We have

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

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

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

We bound the two terms separately. We have

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

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

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

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

$$\Delta_{G1} = O_P \left( p^{1-a_1} \sqrt{\frac{\log p}{T}} + \frac{1}{p^{3(a_1-\frac{1}{2})-c}} + \frac{1}{p^{a_1-ca_2}} \right). \quad (\text{A.7})$$

For  $\Delta_{G3}$ , we have

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

By Lemma A.3, we have

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

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

$$\Delta_{G_3}^{(a)} \leq p^{-1/2} \|\Theta^{-1}\| \|\Omega' \widehat{\mathbf{V}}\|_F^2 \|\widehat{\mathbf{\Gamma}}\| = O_P(p^{11/2-5a_1} \log p/T + 1/p^{5a_1-7/2-2c}).$$

Similarly,  $\Delta_{G_3}^{(b)} = O_P(1/p^{5a_1-7/2-2c})$  because  $\|\Omega' \widetilde{\mathbf{V}}\|_F = O(\sqrt{p} \|\widetilde{\mathbf{V}} - \mathbf{V}\|_{\max}) = O_P(1/p^{3a_1-2-c})$

by Lemma A.2. Then, we obtain

$$\Delta_{G_3} = O_P\left(p^{\frac{11}{2}-5a_1} \frac{\log p}{T} + \frac{1}{p^{5a_1-\frac{7}{2}-2c}}\right).$$

Similarly, we can show that the terms  $\Delta_{G_2}$ ,  $\Delta_{G_4}$ ,  $\Delta_{G_5}$  and  $\Delta_{G_6}$  are dominated by  $\Delta_{G_1}$  and  $\Delta_{G_3}$ . Therefore, we have

$$\Delta_G = O_P\left(p^{1-a_1} \sqrt{\frac{\log p}{T}} + \frac{1}{p^{a_1-ca_2}} + p^{\frac{11}{2}-5a_1} \frac{\log p}{T} + \frac{1}{p^{5a_1-\frac{7}{2}-2c}}\right). \quad (\text{A.8})$$

Similarly, we consider

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

For  $\Delta_{L2}$ , similar to the proof of (A.8), we have

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

We have

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

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

$$\|\widehat{\Phi} - \Phi\|_F^2 = \sum_{j=1}^G \|\widehat{\Phi}^j - \Phi^j\|_F^2 = O_P(p^{1-c} (p^{5(1-a_1)+6c(1-a_2)} \frac{\log p}{T} + \frac{1}{p^{5a_1-3+2c(3a_2-4)}} + \frac{m_p^2}{p^{2c(3a_2-2)}})).$$

Then,  $II$  is of order  $O_P(p^{\frac{5}{2}(1-a_1)+c(\frac{5}{2}-3a_2)} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1-\frac{3}{2}+c(3a_2-\frac{7}{2})} + m_p/p^{3c(a_2-\frac{1}{2})})$  and  $I$  is of smaller order. By Lemma A.4, we have  $III = O_P(p^{\frac{5}{2}(1-a_1)+c(1-a_2)} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2})$ . Thus,  $\Delta_{L2}^{(a)} = O_P(p^{\frac{5}{2}(1-a_1)+c(1-a_2)} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2} + m_p/p^{3c(a_2-\frac{1}{2})})$ .

Similarly, we have

$$\begin{aligned}\Delta_{L2}^{(b)} &\leq p^{-1/2} \left( \|\Psi^{-1/2}(\Phi' \widetilde{\Phi} - \mathbf{I}) \widetilde{\Psi}(\widetilde{\Phi}' \Phi - \mathbf{I}) \Psi^{-1/2}\|_F + 2\|\Psi^{-1/2}(\Phi' \widetilde{\Phi} - \mathbf{I}) \widetilde{\Psi} \Psi^{-1/2}\|_F \right. \\ &\quad \left. + \|(\Psi^{-1/2}(\widetilde{\Psi} - \Psi) \Psi^{-1/2})\|_F \right) =: I' + II' + III'.\end{aligned}$$

By  $\sin \theta$  theorem,  $\|\Phi' \widetilde{\Phi} - \mathbf{I}\| \leq \|\widetilde{\Phi} - \Phi\| \leq \max_j \|\widetilde{\Phi}^j - \Phi^j\| \leq O(m_p/p^{ca_2})$ . 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

$$\Delta_{L2} = O_P \left( p^{\frac{5}{2}(1-a_1)+c(1-a_2)} \sqrt{\frac{\log p}{T}} + \frac{1}{p^{\frac{5}{2}a_1-\frac{3}{2}-2c+ca_2}} + \frac{m_p}{p^{ca_2}} \right). \quad (\text{A.9})$$

For  $\Delta_{L3}$ , we have

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

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

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

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  $\|\Omega' \widetilde{\Phi}\|_F^2 \leq \|\widetilde{\Phi} - \Phi\|_F^2 = \sum_{j=1}^G \|\widetilde{\Phi}^j - \Phi^j\|_F^2 = O(m_p^2/p^{3c(2a_2-1)-1})$ .

Similarly, we can show  $\Delta_{L1}$ ,  $\Delta_{L4}$ ,  $\Delta_{L5}$  and  $\Delta_{L6}$  are dominated by  $\Delta_{L2}$  and  $\Delta_{L3}$ . Therefore, we have

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

Combining the terms  $\Delta_G$ ,  $\Delta_L$  and  $\Delta_S$  together, we complete the proof of (3.5).  $\square$

## A.2 Proof for Section 3.1

We consider (3.6). Similar to the proof of Theorem 2.1 in Fan et al. (2011), we can obtain

$$\|\widehat{\Sigma}_E^T - \Sigma_E\|_2 = O_P(\mu_p \widetilde{\omega}_T^{1-q}) \text{ from (A.2).}$$

Consider (3.7). We have

$$\begin{aligned} \|\widehat{\Sigma}^{\mathcal{T}} - \Sigma\|_{\Sigma} &\leq p^{-1/2} \left( \|\Sigma^{-1/2}(\widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}')\Sigma^{-1/2}\|_F + \|\Sigma^{-1/2}(\widehat{\Sigma}_E^{\mathcal{T}} - \Sigma_E)\Sigma^{-1/2}\|_F \right) \\ &=: \Delta_G + \Delta_E \end{aligned}$$

and

$$\Delta_E = O_P(p^{-1/2}\|\widehat{\Sigma}_E^{\mathcal{T}} - \Sigma_E\|_F) = O_P(\|\widehat{\Sigma}_E^{\mathcal{T}} - \Sigma_E\|_2) = O_P(\mu_p\widetilde{\omega}_T^{1-q}).$$

Combining the terms  $\Delta_E$  and  $\Delta_G$  from (A.8), we have

$$\|\widehat{\Sigma}^{\mathcal{T}} - \Sigma\|_{\Sigma} = O_P \left( \mu_p\widetilde{\omega}_T^{1-q} + p^{\frac{11}{2}-5a_1} \frac{\log p}{T} + \frac{1}{p^{5a_1-\frac{7}{2}-2c}} \right).$$

### A.3 Proofs for Section 3.2

Under the orthogonality condition, we have the following modified rates of convergence. Define  $\mathbf{W} = [\mathbf{B} \quad \mathbf{\Lambda}]$ , which is a  $p \times (k+r)$  loading matrix. Let  $\{\bar{\delta}_i, \bar{v}_i\}_{i=1}^{k+r}$  be the leading eigenvalues and eigenvectors of  $\mathbf{W}\mathbf{W}'$  in the decreasing order and the rest zero.

**Lemma A.5.** *Under the assumptions of Theorem 3.2, we have the following results.*

(i) *By Weyl's theorem, we have*

$$|\delta_i - \bar{\delta}_i| \leq \|\Sigma_u\| \text{ for } i \leq k+r, \quad |\delta_i| \leq \|\Sigma_u\| \text{ for } i > k+r,$$

$|\delta_i - \bar{\delta}_i| \leq \|\Sigma_u\|$  for  $i \leq k+r$  and  $|\delta_i| \leq \|\Sigma_u\|$  for  $i > k+r$ . In addition, for all  $p$ ,  $\bar{\delta}_i/p^{a_1}$  and  $\bar{\delta}_i/p^{ca_2}$  are strictly bigger than zero for  $i \leq k$  and  $k < i \leq k+r$ , respectively.

(ii) *We have*

$$\max_{i \leq k} \|\bar{v}_i - v_i\|_{\infty} = O \left( \frac{m_p}{p^{3(a_1-\frac{1}{2})}} \right),$$

$$\max_{i \leq k} \|\widehat{v}_i - v_i\|_\infty = O_P \left( p^{3(1-a_1)} \sqrt{\frac{\log p}{Tp}} + \frac{m_p}{p^{3(a_1 - \frac{1}{2})}} \right).$$

*Proof.* (ii) Under the orthogonality condition, we have

$$\widehat{\Sigma} = \mathbf{B}\mathbf{B}' + \Lambda\Lambda' + \Sigma_u + (\widehat{\Sigma} - \Sigma) = \mathbf{W}\mathbf{W}' + \Sigma_u + (\widehat{\Sigma} - \Sigma).$$

We can treat  $\mathbf{W}\mathbf{W}'$  as a low rank matrix and the remaining terms as a perturbation matrix.

Then, similar to Lemmas A.2 and A.3, we can show

$$\begin{aligned} \max_{i \leq k} \|\widehat{v}_i - v_i\|_\infty &\leq Cp^{2(1-a_1)} \frac{\|\Sigma_u\|_\infty}{\bar{\gamma}\sqrt{p}}, \\ \max_{i \leq k} \|\widehat{v}_i - v_i\|_\infty &\leq Cp^{2(1-a_1)} \frac{\|\Sigma_u + (\widehat{\Sigma} - \Sigma)\|_\infty}{\bar{\gamma}\sqrt{p}} = O_P \left( \frac{\|\Sigma_u\|_\infty}{p^{3(a_1 - \frac{1}{2})}} + p^{3(1-a_1)} \sqrt{\frac{\log p}{Tp}} \right), \end{aligned}$$

where the eigengap  $\bar{\gamma} = \min\{\bar{\delta}_i - \bar{\delta}_{i+1} : 1 \leq i \leq k\}$ . □

**Lemma A.6.** *Under the assumptions of Theorem 3.2, for  $i \leq r_j$ , we have*

$$\begin{aligned} |\widehat{\kappa}_i^j / \kappa_i^j - 1| &= O_P \left( p^{\frac{5}{2}(1-a_1) + c(1-a_2)} \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{\frac{5a_1}{2} - \frac{3}{2} - c(1-a_2)}} \right), \\ \|\widehat{\eta}_i^j - \eta_i^j\|_\infty &= O_P \left( p^{\frac{5}{2}(1-a_1) + c(\frac{5}{2} - 3a_2)} \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{\frac{5a_1}{2} - \frac{3}{2} + c(3a_2 - \frac{5}{2})}} + \frac{m_p}{p^{3c(a_2 - \frac{1}{2})}} \right). \end{aligned}$$

*Proof.* Similar to the proof of Lemma A.4 using results from Lemma A.5, we can obtain

$$\|\widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)} \sqrt{\log p/T} + m_p/p^{\frac{5a_1}{2} - \frac{3}{2}}). \quad (\text{A.11})$$

Then, we have

$$\|\widehat{\Sigma}_E - \Sigma_E\|_{\max} \leq \|\widehat{\Sigma} - \Sigma\|_{\max} + \|\widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}' - \mathbf{B}\mathbf{B}'\|_{\max}$$

$$= O_P(p^{\frac{5}{2}(1-a_1)}\sqrt{\log p/T} + m_p/p^{\frac{5a_1}{2}-\frac{3}{2}}). \quad (\text{A.12})$$

Therefore, the first statement is followed by (A.12) and the Weyl's theorem. In addition, by Theorem 1 of Fan et al. (2018b) and (A.12), we have

$$\begin{aligned} \|\widehat{\eta}_i^j - \eta_i^j\|_\infty &\leq C p_j^{2(1-a_2)} \frac{\|\Sigma_u^j + (\widehat{\Sigma}_E^j - \Sigma_E^j)\|_\infty}{p_j^{a_2} \sqrt{p_j}} \\ &= O_P \left( p^{\frac{5}{2}(1-a_1)+c(\frac{5}{2}-3a_2)} \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{\frac{5a_1}{2}-\frac{3}{2}+c(3a_2-\frac{5}{2})}} + \frac{m_p}{p^{3c(a_2-\frac{1}{2})}} \right). \end{aligned}$$

□

**Proof of Theorem 3.2.** Using the similar proof of (A.5) and results from Lemmas A.5 and A.6, we can obtain

$$\|\widehat{\Phi}\widehat{\Psi}\widehat{\Phi}' - \Lambda\Lambda'\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\sqrt{\log p/T} + m_p/p^{\frac{5}{2}a_1-\frac{3}{2}-\frac{5}{2}c(1-a_2)} + m_p/\sqrt{p^{c(5a_2-3)}}). \quad (\text{A.13})$$

By Assumption 3.1(iii), (A.11) and (A.13), we have

$$\|\widehat{\Sigma}_u - \Sigma_u\|_{\max} = O_P \left( p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)} \sqrt{\frac{\log p}{T}} + \frac{m_p}{\sqrt{p^{5a_1-3-5c(1-a_2)}}} + \frac{m_p}{\sqrt{p^{c(5a_2-3)}}} \right).$$

Therefore,  $\|\widehat{\Sigma}_u^{\mathcal{D}} - \Sigma_u\|_{\max} = O_P(\tau + \omega_{T,2}) = O_P(\omega_{T,2})$ , when  $\tau$  is chosen as the same order of  $\omega_{T,2} = p^{\frac{5}{2}(1-a_1)+\frac{5}{2}c(1-a_2)}\sqrt{\log p/T} + m_p/\sqrt{p^{5a_1-3-5c(1-a_2)}} + m_p/\sqrt{p^{c(5a_2-3)}}$ . Similar to the proofs of Theorem 3.1, we can obtain (3.8) – (3.11).

Consider (3.12). First, we have

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

and

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

By Lemma A.5, we have

$$\begin{aligned} \|\mathbf{V}'\widehat{\mathbf{V}} - \mathbf{I}\|_F &= \|\mathbf{V}'(\widehat{\mathbf{V}} - \mathbf{V})\|_F \leq \|\widehat{\mathbf{V}} - \mathbf{V}\|_F = O_P(p^{3(1-a_1)} \sqrt{\log p/T} + m_p/p^{3a_1-2}), \\ \|\mathbf{V}'\widetilde{\mathbf{V}} - \mathbf{I}\|_F &= \|\mathbf{V}'(\widetilde{\mathbf{V}} - \mathbf{V})\|_F \leq \|\widetilde{\mathbf{V}} - \mathbf{V}\|_F = O(m_p/p^{3a_1-2}). \end{aligned}$$

Then, similar to the proof of (A.8), we can obtain

$$\Delta_{G'} = O_P \left( p^{1-a_1} \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{a_1}} + p^{\frac{11}{2}-5a_1} \frac{\log p}{T} + \frac{m_p^2}{p^{5a_1-\frac{7}{2}}} \right). \quad (\text{A.14})$$

By Lemma A.6, we have

$$\begin{aligned} \|\widehat{\Phi} - \Phi\|_F^2 &= \sum_{j=1}^J \|\widehat{\Phi}^j - \Phi^j\|_F^2 \\ &= O_P \left( p^{1-c} \left( p^{5(1-a_1)+6c(1-a_2)} \frac{\log p}{T} + \frac{m_p^2}{p^{5a_1-3-6c(1-a_2)}} + \frac{m_p^2}{p^{2c(3a_2-2)}} \right) \right), \\ \|\widetilde{\Phi} - \Phi\|_F^2 &= \sum_{j=1}^J \|\widetilde{\Phi}^j - \Phi^j\|_F^2 = O(m_p^2/p^{2c(3a_2-2)-(1-c)}). \end{aligned}$$

Then, similar to the proof of (A.10), we can obtain

$$\begin{aligned} \Delta_{L'} &= O_P \left( p^{\frac{5}{2}(1-a_1)+c(1-a_2)} \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{\frac{5}{2}a_1-\frac{3}{2}-c(1-a_2)}} + \frac{m_p}{p^{ca_2}} \right. \\ &\quad \left. + p^{\frac{11}{2}-5a_1+5c(1-a_2)} \frac{\log p}{T} + \frac{m_p^2}{p^{5a_1-\frac{7}{2}-5c(1-a_2)}} + \frac{m_p^2}{p^{5ca_2-3c-\frac{1}{2}}} \right). \end{aligned}$$

Combining the terms  $\Delta_{G'}$ ,  $\Delta_{L'}$  and  $\Delta_{S'}$  together, we complete the proof of (3.12).

**Proofs for (3.13) and (3.14).** We consider (3.13). Let  $\mathbf{W} = (\tilde{w}_1, \dots, \tilde{w}_{k+r})$ . Then,  $\bar{\delta}_i = \|\tilde{w}_i\|^2 \asymp p$  for  $i \leq k$ ,  $\bar{\delta}_i = \|\tilde{w}_i\|^2 \asymp p^c$  for  $k < i \leq k+r$  by Lemma A.5(i), and  $\bar{v}_i = \tilde{w}_i/\|\tilde{w}_i\|$ . Hence,  $\|\bar{v}_i\|_\infty \leq \|\mathbf{W}\|_{\max}/\|\tilde{w}_i\| \leq C/\sqrt{p^c}$  for  $i \leq k+r$ . In addition, for

$\tilde{\mathbf{V}}_2 = (\tilde{v}_1, \dots, \tilde{v}_{k+r})$ , the coherence  $\mu(\tilde{\mathbf{V}}_2) = \frac{p}{k+r} \max_i \sum_{j=1}^{k+r} \tilde{\mathbf{V}}_{2,ij}^2 \leq Cp^{1-c}$ , where  $\tilde{\mathbf{V}}_{2,ij}$  is the  $(i, j)$  entry of  $\tilde{\mathbf{V}}_2$ . Thus, by Theorem 1 of Fan et al. (2018b) and  $r \asymp p^{1-c}$ , we have

$$\begin{aligned} \max_{i \leq k+r} \|\tilde{v}_i - v_i\|_\infty &\leq C(k+r)^4 p^{2(1-c)} \frac{\|\Sigma_u\|_\infty}{p^c \sqrt{p}} = O\left(\frac{m_p}{p^{7c - \frac{11}{2}}}\right), \\ \max_{i \leq k+r} \|\hat{v}_i - v_i\|_\infty &\leq C(k+r)^4 p^{2(1-c)} \frac{\|\Sigma_u + (\hat{\Sigma} - \Sigma)\|_\infty}{p^c \sqrt{p}} = O_P\left(\frac{m_p}{p^{7c - \frac{11}{2}}} + p^{\frac{13}{2} - 7c} \sqrt{\frac{\log p}{T}}\right). \end{aligned}$$

By the similar argument of (A.1), we can obtain  $\|\hat{\mathbf{V}}_2 \hat{\Gamma}_2 \hat{\mathbf{V}}_2' - \mathbf{W}\mathbf{W}'\|_{\max} = O_P(p^{7(1-c)} \sqrt{\log p/T} + m_p/p^{7c-6})$ . Hence, we have

$$\|\hat{\Sigma}_{u,2} - \Sigma_u\|_{\max} \leq \|\hat{\Sigma} - \Sigma\|_{\max} + \|\hat{\mathbf{V}}_2 \hat{\Gamma}_2 \hat{\mathbf{V}}_2' - \mathbf{W}\mathbf{W}'\|_{\max} = O_P(p^{7(1-c)} \sqrt{\log p/T} + m_p/p^{7c-6}).$$

Similar to the proof of (3.2), we can obtain  $\|\hat{\Sigma}_{u,2}^\mathcal{T} - \Sigma_u\|_2 = O_P(m_p \tilde{\omega}_{T,2}^{1-q})$ , where  $\tilde{\omega}_{T,2} = p^{7(1-c)} \sqrt{\log p/T} + m_p/p^{7c-6}$ .

Consider (3.14). We have

$$\begin{aligned} \|\hat{\Sigma}_2^\mathcal{T} - \Sigma\|_\Sigma &\leq p^{-1/2} \left( \|\Sigma^{-1/2} (\hat{\mathbf{V}}_2 \hat{\Gamma}_2 \hat{\mathbf{V}}_2' - \mathbf{W}\mathbf{W}') \Sigma^{-1/2}\|_F + \|\Sigma^{-1/2} (\hat{\Sigma}_{u,2}^\mathcal{T} - \Sigma_u) \Sigma^{-1/2}\|_F \right) \\ &=: \Delta_{G_2} + \Delta_{S_2} \end{aligned}$$

and

$$\Delta_{S_2} = O_P(p^{-1/2} \|\hat{\Sigma}_{u,2}^\mathcal{T} - \Sigma_u\|_F) = O_P(\|\hat{\Sigma}_{u,2}^\mathcal{T} - \Sigma_u\|_2) = O_P(m_p \tilde{\omega}_{T,2}^{1-q}).$$

Note that  $\|\hat{\mathbf{V}}_2 - \mathbf{V}_2\|_F \leq \sqrt{p^{2-c}} \|\hat{v}_i - v_i\|_\infty = O_P(m_p / \sqrt{p^{15c-13}} + p^{\frac{15}{2}(1-c)} \sqrt{\log p/T})$ , where  $\mathbf{V}_2 = (v_1, \dots, v_{k+r})$  is the leading eigenvectors of covariance matrix  $\Sigma$ . Similar to the proof of (A.8), we can obtain

$$\Delta_{G_2} = O_P\left( (p^{7-\frac{15}{2}c} + p^{1-c}) \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^c} + \frac{m_p}{p^{\frac{15}{2}c-6}} + p^{15(1-c)+\frac{1}{2}} \frac{\log p}{T} + \frac{m_p^2}{p^{15c-\frac{27}{2}}} \right).$$

Combining the terms  $\Delta_{G_2}$  and  $\Delta_{S_2}$ , we have

$$\|\widehat{\Sigma}_2^{\mathcal{T}} - \Sigma\|_{\Sigma} = O_P \left( m_p \widetilde{\omega}_{T,2}^{1-q} + p^{15(1-c)+\frac{1}{2}} \frac{\log p}{T} + \frac{m_p^2}{p^{15c-\frac{27}{2}}} \right).$$

#### A.4 Proof of Theorem 3.3

**Proof of Theorem 3.3.** We consider (3.15). By (3.10), for each group  $j$ , we have

$$\|(\widehat{\Sigma}^{\mathcal{D}})^j - \Sigma^j\|_{\max} \leq \|\widehat{\Sigma}^{\mathcal{D}} - \Sigma\|_{\max} = O_P(\omega_{T,2}).$$

Consider (3.16). Since  $\lambda_{\min}(\Sigma_u^j) > c_1$  and  $m_{p_j} \omega_{T,2}^{1-q} = o(1)$ , the minimum eigenvalue of  $(\widehat{\Sigma}_u^{\mathcal{D}})^j$  is strictly bigger than 0 with probability approaching 1. Then, we have  $\|((\widehat{\Sigma}_u^{\mathcal{D}})^j)^{-1} - (\Sigma_u^j)^{-1}\|_2 \leq \lambda_{\min}(\Sigma_u^j)^{-1} \|(\widehat{\Sigma}_u^{\mathcal{D}})^j - \Sigma_u^j\|_2 \lambda_{\min}((\widehat{\Sigma}_u^{\mathcal{D}})^j)^{-1} = O_P(m_{p_j} \omega_{T,2}^{1-q})$ . Similar to the proofs of (3.4), we can show the statement.

Consider (3.17). We denote  $(\mathbf{A})^j$  the  $j$ th diagonal block for a matrix  $\mathbf{A}$ . We have  $B^j B^{j'} = (\mathbf{B}\mathbf{B}')^j = (\widetilde{\mathbf{V}}\widetilde{\Gamma}\widetilde{\mathbf{V}}')^j = \widetilde{\mathbf{V}}^j \widetilde{\Gamma} \widetilde{\mathbf{V}}^{j'}$ , where  $\widetilde{\mathbf{V}}^j$  is the  $j$ th  $p_j \times k$  submatrix of  $\widetilde{\mathbf{V}}$ . Similarly,  $(\widehat{\mathbf{V}}\widehat{\Gamma}\widehat{\mathbf{V}}')^j = \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'}$ , where  $\widehat{\mathbf{V}}^j$  is the  $j$ th  $p_j \times k$  submatrix of  $\widehat{\mathbf{V}}$ . For each group  $j$ , we have

$$(\widehat{\Sigma}^{\mathcal{D}})^j = \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} + \widehat{\Phi}^j \widehat{\Psi}^j \widehat{\Phi}^{j'} + (\widehat{\Sigma}_u^{\mathcal{D}})^j.$$

By SVD, we have  $\Sigma^j = \mathbf{V}_{p_j} \Gamma_{p_j} \mathbf{V}_{p_j}'$ , where  $\mathbf{V}_{p_j} = (\mathbf{V}^j, \Phi^j, \Omega^j)$  and  $\Gamma_{p_j} = \text{diag}(\Gamma, \Psi^j, \Theta^j)$ .

Then, we have

$$\begin{aligned} \|(\widehat{\Sigma}^{\mathcal{D}})^j - \Sigma^j\|_{\Sigma^j} &\leq (p^j)^{-1/2} \left( \left\| (\Sigma^j)^{-\frac{1}{2}} \left( \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} - B^j B^{j'} \right) (\Sigma^j)^{-\frac{1}{2}} \right\|_F \right. \\ &\quad \left. + \left\| (\Sigma^j)^{-\frac{1}{2}} \left( \widehat{\Phi}^j \widehat{\Psi}^j \widehat{\Phi}^{j'} - \Lambda^j \Lambda^{j'} \right) (\Sigma^j)^{-\frac{1}{2}} \right\|_F + \left\| (\Sigma^j)^{-\frac{1}{2}} \left( (\widehat{\Sigma}_u^{\mathcal{D}})^j - \Sigma_u^j \right) (\Sigma^j)^{-\frac{1}{2}} \right\|_F \right) \\ &=: \Delta_g + \Delta_l + \Delta_s. \end{aligned}$$

Similar to the proofs of Theorem 2.1 in Fan et al. (2011), we can show  $\|(\widehat{\Sigma}_u^{\mathcal{D}})^j - \Sigma_u^j\|_2 =$

$O_P(m_{p_j} \omega_{T,2}^{1-q})$ . Then, we have

$$\Delta_s = O_P(p_j^{-1/2} \|(\widehat{\Sigma}_u^{\mathcal{D}})^j - \Sigma_u^j\|_F) = O_P(\|(\widehat{\Sigma}_u^{\mathcal{D}})^j - \Sigma_u^j\|_2) = O_P(m_{p_j} \omega_{T,2}^{1-q}).$$

We have

$$\begin{aligned} \Delta_g &\leq p_j^{-1/2} \left( \left\| \Gamma^{-1/2} \mathbf{V}^{j'} \left( \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} - B^j B^{j'} \right) \mathbf{V}^j \Gamma^{-1/2} \right\|_F \right. \\ &\quad + \left\| \Psi^{j-1/2} \Phi^{j'} \left( \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} - B^j B^{j'} \right) \Phi^j \Psi^{j-1/2} \right\|_F + \left\| \Theta^{j-1/2} \Omega^{j'} \left( \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} - B^j B^{j'} \right) \Omega^j \Theta^{j-1/2} \right\|_F \\ &\quad + 2 \left\| \Gamma^{-1/2} \mathbf{V}^{j'} \left( \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} - B^j B^{j'} \right) \Phi^j \Psi^{j-1/2} \right\|_F + 2 \left\| \Gamma^{-1/2} \mathbf{V}^{j'} \left( \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} - B^j B^{j'} \right) \Omega^j \Theta^{j-1/2} \right\|_F \\ &\quad \left. + 2 \left\| \Psi^{j-1/2} \Phi^{j'} \left( \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} - B^j B^{j'} \right) \Omega^j \Theta^{j-1/2} \right\|_F \right) \\ &=: \Delta_{g1} + \Delta_{g2} + \Delta_{g3} + 2\Delta_{g4} + 2\Delta_{g5} + 2\Delta_{g6}. \end{aligned}$$

Using the similar proof of (A.7) and results from Lemma A.5, we can obtain

$$\Delta_{g1} = O_P \left( p^{1-a_1} \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{a_1}} \right).$$

For  $\Delta_{g3}$ , we have

$$p_j^{-1/2} \left( \left\| \Theta^{j-1/2} \Omega^{j'} \widehat{\mathbf{V}}^j \widehat{\Gamma} \widehat{\mathbf{V}}^{j'} \Omega^j \Theta^{j-1/2} \right\|_F + \left\| \Theta^{j-1/2} \Omega^{j'} \widetilde{\mathbf{V}}^j \widetilde{\Gamma} \widetilde{\mathbf{V}}^{j'} \Omega^j \Theta^{j-1/2} \right\|_F \right) =: \Delta_{g3}^{(a)} + \Delta_{g3}^{(b)}.$$

By Lemma A.5, we have

$$\|\widehat{\mathbf{V}}^j - \mathbf{V}^j\|_F = O_P(\sqrt{p^c} \|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max}) = O_P \left( \sqrt{p^c} (p^{3(1-a_1)} \sqrt{\log p / T p} + m_p / p^{3(a_1 - \frac{1}{2})}) \right).$$

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

$$\Delta_{g3}^{(a)} \leq p_j^{-1/2} \|\Theta^{j-1}\| \|\Omega^{j'} (\widehat{\mathbf{V}}^j - \mathbf{V}^j)\|_F^2 \|\widehat{\Gamma}\| = O_P(p^{5(1-a_1)+c/2} \log p / T + m_p^2 / p^{5a_1-3-c/2}).$$

Similarly, since  $\|\tilde{\mathbf{V}}^j - \mathbf{V}^j\|_F = O_P(\sqrt{p^c}\|\tilde{\mathbf{V}} - \mathbf{V}\|_{\max}) = O_P(m_p/p^{3(a_1-1/2)-c/2})$  by Lemma A.5,  $\Delta_{g3}^{(b)} = O_P(m_p^2/p^{5a_1-3-c/2})$ . Then, we obtain

$$\Delta_{g3} = O_P\left(p^{5(1-a_1)+\frac{c}{2}}\frac{\log p}{T} + \frac{m_p^2}{p^{5a_1-3-\frac{c}{2}}}\right).$$

Similarly, we can show that the terms  $\Delta_{g2}$ ,  $\Delta_{g4}$ ,  $\Delta_{g5}$  and  $\Delta_{g6}$  are dominated by  $\Delta_{g1}$  and  $\Delta_{g3}$ . Thus, we have

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

We have

$$\begin{aligned} \Delta_l &\leq p_j^{-1/2}\left(\left\|\mathbf{\Gamma}^{-1/2}\mathbf{V}^{j'}(\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Lambda^j\Lambda^{j'})\mathbf{V}^j\mathbf{\Gamma}^{-1/2}\right\|_F\right. \\ &\quad + \left\|\Psi^{j-1/2}\Phi^{j'}(\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Lambda^j\Lambda^{j'})\Phi^j\Psi^{j-1/2}\right\|_F + \left\|\Theta^{j-1/2}\Omega^{j'}(\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Lambda^j\Lambda^{j'})\Omega^j\Theta^{j-1/2}\right\|_F \\ &\quad + 2\left\|\mathbf{\Gamma}^{-1/2}\mathbf{V}^{j'}(\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Lambda^j\Lambda^{j'})\Phi^j\Psi^{j-1/2}\right\|_F + 2\left\|\mathbf{\Gamma}^{-1/2}\mathbf{V}^{j'}(\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Lambda^j\Lambda^{j'})\Omega^j\Theta^{j-1/2}\right\|_F \\ &\quad \left. + 2\left\|\Psi^{j-1/2}\Phi^{j'}(\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'} - \Lambda^j\Lambda^{j'})\Omega^j\Theta^{j-1/2}\right\|_F\right) \\ &=: \Delta_{l1} + \Delta_{l2} + \Delta_{l3} + 2\Delta_{l4} + 2\Delta_{l5} + 2\Delta_{l6}. \end{aligned}$$

For  $\Delta_{l2}$ , similar to the proof of (A.9) using results from Lemma A.6, we can obtain

$$\Delta_{l2} = O_P\left(p^{\frac{5}{2}(1-a_1)+c(1-a_2)}\sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{\frac{5}{2}a_1-\frac{3}{2}-c(1-a_2)}} + \frac{m_p}{p^{ca_2}}\right).$$

For  $\Delta_{l3}$ , we have

$$\Delta_{l3} \leq p_j^{-1/2}\left(\left\|\Theta^{j-1/2}\Omega^{j'}\widehat{\Phi}^j\widehat{\Psi}^j\widehat{\Phi}^{j'}\Omega^j\Theta^{j-1/2}\right\|_F + \left\|\Theta^{j-1/2}\Omega^{j'}\tilde{\Phi}^j\tilde{\Psi}^j\tilde{\Phi}^{j'}\Omega^j\Theta^{j-1/2}\right\|_F\right) =: \Delta_{l3}^{(a)} + \Delta_{l3}^{(b)}.$$

By Lemma A.6, we have  $\|\widehat{\Phi}^j - \Phi^j\|_F \leq \sqrt{p_j r_j}\|\widehat{\Phi}^j - \Phi^j\|_{\max} = O_P(p^{\frac{5}{2}(1-a_1)+3c(1-a_2)}\sqrt{\log p/T} +$

$m_p/p^{\frac{5a_1}{2}-\frac{3}{2}-3c(1-a_2)} + m_p/p^{c(3a_2-2)}$ ) and  $\|\widehat{\Psi}^j\| = O_P(p^{ca_2})$ , we have

$$\begin{aligned}\Delta_{l_3}^{(a)} &\leq p_j^{-1/2} \|\Theta^{j-1}\| \|\Omega^{j'}(\widehat{\Phi}^j - \Phi^j)\|_F^2 \|\widehat{\Psi}^j\| \\ &= O_P\left(p^{5(1-a_1)+c(\frac{11}{2}-5a_2)} \frac{\log p}{T} + \frac{m_p^2}{p^{5a_1-3+c(5a_2-\frac{11}{2})}} + \frac{m_p^2}{p^{c(5a_2-\frac{7}{2})}}\right).\end{aligned}$$

Similarly, by Lemma A.2,  $\|\widetilde{\Phi}^j - \Phi^j\|_F = O_P(\sqrt{p^c} \|\widetilde{\Phi}^j - \Phi^j\|_{\max}) = O_P(m_p/p^{c(3a_2-2)})$ , then  $\Delta_{l_3}^{(b)} = O_P(m_p^2/p^{c(5a_2-\frac{7}{2})})$ . Similarly, we can show  $\Delta_{l_1}$ ,  $\Delta_{l_4}$ ,  $\Delta_{l_5}$  and  $\Delta_{l_6}$  are dominated by  $\Delta_{l_2}$  and  $\Delta_{l_3}$ . Thus, we have

$$\begin{aligned}\Delta_l &= O_P\left(p^{\frac{5}{2}(1-a_1)+c(1-a_2)} \sqrt{\frac{\log p}{T}} + \frac{m_p}{p^{\frac{5}{2}a_1-\frac{3}{2}-c(1-a_2)}} + \frac{m_p}{p^{ca_2}}\right. \\ &\quad \left.+ p^{5(1-a_1)+c(\frac{11}{2}-5a_2)} \frac{\log p}{T} + \frac{m_p^2}{p^{5a_1-3+c(5a_2-\frac{11}{2})}} + \frac{m_p^2}{p^{c(5a_2-\frac{7}{2})}}\right).\end{aligned}$$

Combining the terms  $\Delta_g$ ,  $\Delta_l$  and  $\Delta_s$  together, we complete the proof of (3.17).  $\square$

## A.5 Proof of Theorem 3.4

**Lemma A.7.** *Under the assumptions of Theorem 3.4, for  $i = 1, \dots, [dm] - 2(k+r)$ , there exist constants  $\underline{c}, \bar{c} > 0$  such that*

$$\underline{c} + o_P(1) \leq m\widehat{\delta}_{k+r+i}/p \leq \bar{c} + o_P(1).$$

*Proof.* Similar to the proofs of Lemma A.9 in Ahn and Horenstein (2013), we can show the statement.  $\square$

**Lemma A.8.** *Under the multi-level factor model and assumptions of Theorem 3.4, for  $i = 1, \dots, k$ , we have*

$$\widehat{\delta}_i = \bar{\delta}_i + O_P\left(p\sqrt{\frac{\log p}{T}}\right) + O_P(p^{ca_2}).$$

In addition, for  $i = k + 1, \dots, k + r$ , we have

$$\widehat{\delta}_i = \bar{\delta}_i + O_P \left( p^{\frac{5}{2}(1-a_1)+c} \sqrt{\frac{\log p}{T}} + \frac{1}{p^{\frac{5}{2}a_1 - \frac{3}{2} - 2c}} \right) + O_P(m_p).$$

*Proof.* By Lemmas A.1 and A.3, the first statement is showed. The order of  $|\widehat{\delta}_i - \bar{\delta}_i|$  for  $i = k + 1, \dots, k + r$  is the same as that of  $|\widehat{\kappa}_i^j - \bar{\kappa}_i^j|$  for  $i = 1, \dots, r_j$  and each group  $j$ . Then the second statement is showed from Lemmas A.1 and A.4.  $\square$

**Proof of Theorem 3.4.** (i) When  $k > 0$  and  $r > 0$ , by Lemmas A.1 and A.8,  $\widehat{\delta}_i/\widehat{\delta}_{i+1} = \bar{\delta}_i/\bar{\delta}_{i+1} + o_P(1) = O_P(1)$  for  $i = 1, 2, \dots, k - 1, k + 1, \dots, k + r - 1$ . In addition, we have

$$\frac{\widehat{\delta}_k}{\widehat{\delta}_{k+1}} = \frac{\bar{\delta}_k + O_P \left( p\sqrt{\log p/T} \right) + O_P(p^{ca_2})}{\bar{\delta}_{k+1} + O_P \left( p^{\frac{5}{2}(1-a_1)+c} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} - 2c} \right) + O_P(m_p)} = O_P(p^{a_1-ca_2}).$$

By Lemmas A.7 and A.8, we have

$$\frac{\widehat{\delta}_{k+r}}{\widehat{\delta}_{k+r+1}} \geq \frac{\bar{\delta}_{k+r} + O_P \left( p^{\frac{5}{2}(1-a_1)+c} \sqrt{\log p/T} + 1/p^{\frac{5}{2}a_1 - \frac{3}{2} - 2c} \right) + O_P(m_p)}{p[\bar{c} + o_P(1)]/m} = O_P \left( \frac{m}{p^{1-ca_2}} \right),$$

which diverges to  $\infty$  if  $p^{1-ca_2} \ll T$ . By Lemma A.7, for  $i = 1, \dots, [dm] - 2(k + r) - 1$ ,  $\widehat{\delta}_{k+r+i}/\widehat{\delta}_{k+r+i+1} \leq (\bar{c} + o_P(1))/(\underline{c} + o_P(1))$ .

When  $k = 0$  and  $r > 0$ , for  $i \leq r$ , we have  $|\bar{\delta}_i - \delta_i| \leq \|\Sigma_u^j\| = O(m_p)$  and  $|\widehat{\delta}_i - \delta_i| \leq p^c \|\widehat{\Sigma}^j - \Sigma^j\|_{\max} = O_P(p^c \sqrt{\log p/T})$ . Then,  $\widehat{\delta}_i/\widehat{\delta}_{i+1} = \bar{\delta}_i/\bar{\delta}_{i+1} + o_P(1) = O_P(1)$  for  $i = 1, 2, \dots, r - 1$ .

In addition, by Lemmas A.7, we have

$$\frac{\widehat{\delta}_r}{\widehat{\delta}_{r+1}} \geq \frac{\bar{\delta}_r + O_P \left( p^c \sqrt{\log p/T} \right) + O_P(m_p)}{p[\bar{c} + o_P(1)]/m} = O_P \left( \frac{m}{p^{1-ca_2}} \right),$$

which diverges to  $\infty$  if  $p^{1-ca_2} \ll T$ . By Lemma A.7, for  $i = 1, \dots, [dm] - 2r - 1$ ,  $\widehat{\delta}_{r+i}/\widehat{\delta}_{r+i+1} \leq (\bar{c} + o_P(1))/(\underline{c} + o_P(1))$ .

When  $k > 0$  and  $r = 0$ , for  $i \leq k$ , we have  $|\bar{\delta}_i - \delta_i| \leq \|\Sigma_u\| = O(m_p)$  and  $|\widehat{\delta}_i - \delta_i| =$

$O_P(p\sqrt{\log p/T})$ . Then,  $\widehat{\delta}_i/\widehat{\delta}_{i+1} = O_P(1)$  for  $i = 1, 2, \dots, k-1$ . In addition, we have

$$\frac{\widehat{\delta}_k}{\widehat{\delta}_{k+1}} \geq \frac{\bar{\delta}_k + O_P\left(p\sqrt{\log p/T}\right) + O_P(m_p)}{p[\bar{c} + o_P(1)]/m} = O_P\left(\frac{m}{p^{1-a_1}}\right),$$

which diverges to  $\infty$  if  $p^{1-a_1} \ll T$ . By Lemma A.7, for  $i = 1, \dots, [dm] - 2k - 1$ ,  $\widehat{\delta}_{k+i}/\widehat{\delta}_{k+i+1} \leq (\bar{c} + o_P(1))/(\underline{c} + o_P(1))$ . These results show the consistency of model selection.

(ii) Under the multi-level factor model (i.e.,  $k > 0$  and  $r > 0$ ), the above results imply that  $\hat{k}$  is consistent by using the fact that  $k < k + r$ .  $\square$