Network Assisted Approximate Factor Model Estimation

Yuzhou Zhao, Xinyan Fan* and Bo Zhang

Center for Applied Statistics and School of Statistics, Renmin University of China

Supplementary Material

The Supplementary Material consists of ten sections (S.1–S.10). Section S.1 provides a more general form of Theorem 3. Section S.2 introduces some useful notations and lemmas that are used to prove the theoretical properties in Section 3. Sections S.3–S.7 present the proofs of Theorems 1, 2, S.1 and 3, 4, and Proposition 1, respectively. Section S.8 provides additional algorithmic details. Section S.9 details the comparison methods. Section S.10 presents additional simulation results.

S.1 More General Theoretical Results

In this section, we present a more general result regarding the convergence rates of $\widehat{B}$ and $\widehat{\Sigma}_Y$.

Theorem S.1. *Suppose Assumptions 1–3 and 5 hold. If* $[\log(p)]^{2/r_m - 1} \ll T$*, and* $\lambda \gg d_{p,T}$ *, we have*

$$\min_{OO^\top = O^\top O = I_r} p^{-1} ||\widehat{B}O - B_0||_F^2 = O_p(\theta_{p,T} \log(p)/T + 1/(pv_p^2)),$$

*Corresponding author.

and

$$||\widehat{\Sigma}_Y - \Sigma_{Y_0}||_{\Sigma_{Y_0}} = O_p(\sqrt{p}\theta_{p,T}\log(p)/T + v_p^{-2}/\sqrt{p} + \Delta_{\Sigma_e}),$$

where

$$\Delta_{\Sigma_e} = \left[(\rho_{p,T}K_T + \log(p) + p\sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)} + \rho_{p,T}^2 D_p)/p\right]^{1/2},$$

and

$$p\theta_{p,T} = (\log(p) + p\sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)})/\rho_{p,T} + K_T$$
$$+ \left[(p + D_p)(\rho_{p,T}K_T + \log(p) + p\sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)} + \rho_{p,T}^2 D_p)\right]^{1/2}.$$

This theorem establishes the convergence rate of our estimators when the λ is sufficiently large. It is important to emphasize that, under Assumptions 3 and 5, $\theta_{p,T} = o(1)$, which implies that the convergence rate of $\min_{OO^\top = O^\top O = I_r} p^{-1}||\widehat{B}O - B_0||_F^2$ exceeds the rate $O_p(\log(p)/T)$ when p is sufficiently large. Furthermore, Theorem 3 can be viewed as a special case of Theorem S.1. The proof of Theorem S.1 is provided in Section S.5.

S.2 Some Useful Notations and Lemmas

To simplify the proofs, we first introduce some notations following Bai and Liao (2016). Define

$$Q_1(\Sigma_e) = \frac{1}{p}\log|\Sigma_e| + \frac{1}{p}tr(S_e\Sigma_e^{-1}) + \frac{\rho_{p,T}}{p}\sum_{i\neq j}|\Sigma_{e,ij}|,$$

$$\begin{aligned}Q_2(B,\Sigma_e) &= \frac{1}{p}tr(B_0^\top \Sigma_e^{-1} B_0 - B_0^\top \Sigma_e^{-1} B(B^\top \Sigma_e^{-1} B)^{-1} B^\top \Sigma_e^{-1} B_0)\\ &= \frac{1}{p}||(I_p - \Sigma_e^{-1/2} B(B^\top \Sigma_e^{-1} B)^{-1} B^\top \Sigma_e^{-1/2})\Sigma_e^{-1/2} B_0||_F^2,\end{aligned}$$

and

$$\begin{aligned}Q_3(B,\Sigma_e) =& \frac{1}{p}\log|BB^\top + \Sigma_e| + \frac{1}{p}tr(S_y(BB^\top + \Sigma_e)^{-1})\\ &- \frac{1}{p}\log|\Sigma_e| - \frac{1}{p}tr(S_e\Sigma_e^{-1}) - Q_2(B,\Sigma_e),\end{aligned}$$

where $S_e = \mathcal{E}\mathcal{E}^\top/T$, $\mathcal{E} = (e_1,\ldots,e_T)$. Then

$$\begin{aligned}\frac{1}{p}L(B,\Sigma_e,\Omega,\alpha) =& \frac{1}{p}\log|BB^\top + \Sigma_e| + \frac{1}{p}tr(S_y(BB^\top + \Sigma_e)^{-1})\\ &+ \frac{\rho_{p,T}}{p}\sum_{i\neq j}|\Sigma_{e,ij}| + \frac{\lambda}{pT}L_A(B,\Omega,\alpha)\\ =& Q_1(\Sigma_e) + Q_2(B,\Sigma_e) + Q_3(B,\Sigma_e) + \frac{\lambda}{pT}L_A(B,\Omega,\alpha).\end{aligned}$$

Denote $P_X = X(X^\top X)^{-1}X^\top$ as the projection matrix of X,for any full rank matrix X. Let $x^+ = \max(x,0)$ and $x^- = -\min(x,0)$ for any $x \in \mathbb{R}$. Remind that $d_{p,T} = \max\{\log(p)T/p, \sqrt{\log(p)T}\}$. Denote $X_n = O_p(a_n)$ for matrix X_n if $||X_n||_2 = O_p(a_n)$. It is easy to verify that $O_p(a_n)O_p(b_n) = O_p(a_nb_n)$ and $O_p(a_n) + O_p(b_n) = O_p(a_n + b_n)$.

To prove the theorems, we next introduce the following ten useful lemmas. Lemmas 1(i),(ii), 2, 3, 8, and 9, below are directly modified from Lemmas A.1, A.2, B.1 in Bai and Liao (2016), Lemma A.1 in Bai and Li (2012) and Lemma 28 in Ma et al. (2020), respectively. We only present the proofs of Lemmas 4–7 and 10.

Lemma 1. *Under Assumption 1 and assume that* $[\log(p)]^{2/r_m-1} \ll T$

(i) $\sup_{i,j\leq p}\left|\frac{1}{T}\sum_{t=1}^{T} e_{it}e_{jt} - Ee_{it}e_{jt}\right| = O_p(\sqrt{\log(p)/T})$.

(ii) $\sup_{i\leq r,j\leq p}\left|\frac{1}{T}\sum_{t=1}^{T} f_{it}e_{jt}\right| = O_p(\sqrt{\log(p)/T})$

Lemma 2. *Under Assumption 1 and assume that* $[\log(p)]^{2/r_m-1} \ll T$*,*

$$\sup_{(B,\Sigma_e,\Omega,\alpha)\in\Xi_\delta} |Q_3(B,\Sigma_e)| = O_p(\frac{\log(p)}{p} + \sqrt{\frac{\log(p)}{T}}).$$

Lemma 3. *Denote* $\Delta = \widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1}$*, and remind* $K_T = \sum_{(i,j)\in J_L} |\Sigma_{e0,ij}|$*. For all large enough* p *and* T*, there exists a constant* c_4 *such that*

$$\begin{aligned} pQ_1(\widehat{\Sigma}_e) - pQ_1(\Sigma_{e0}) \geq \frac{1}{2}\rho_{p,T} \sum_{(i,j)\in J_L} \left|\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}\right| + c_4\|\Delta\|_F^2 - 2\rho_{p,T}K_T \\ - \left(O_p\left(\sqrt{\frac{\log(p)}{T}}\right)\sqrt{p+D_p} + \rho_{p,T}\sqrt{D_p}\right)\|\Delta\|_F. \end{aligned}$$

Lemma 4. *Under Assumptions 1-3,4(1), and 5, for* $[\log(p)]^{2/r_m-1} \ll T$ *and* $\lambda \gg \min\{v_p^{2\varepsilon}, \log^{-\varepsilon/(1+\varepsilon)}(p)\}d_{p,T}$*, we have*

$$p^{-1}B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})B_0 = o_p(1).$$

Proof: We prove Lemma 4 following Bai and Liao (2016). Denote $\Delta_1 = \widehat{\Sigma}_e - \Sigma_{e0}$, $B_0^\top\Sigma_{e0}^{-1} = (v_1,\ldots,v_p)$, and $B_0^\top\widehat{\Sigma}_e^{-1} = (u_1,\ldots,u_p)$. According to Assumption 1, $||\Sigma_{e0}^{-1}||_1$ and $||\widehat{\Sigma}_e^{-1}||_1$ are bounded. Then we have $\sup_{i\leq p}||v_i||_2 =$

$O(1)$ and $\sup_{i\le p} ||u_i||_2 = O_p(1)$. After simple calculation, we have

$$B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})B_0 = -B_0^\top \widehat{\Sigma}_e^{-1}\Delta_1\Sigma_{e0}^{-1}B_0$$
$$= -(\sum_{(i,j)\in J_L} u_i v_j^\top \Delta_{1,ij} + \sum_{(i,j)\in J_U} u_i v_j^\top \Delta_{1,ij}).$$

Notice that $\sup_{i,j\le p} \sigma_1(u_i v_j^\top) \le \sup_{i\le p} ||u_i||_2 \sup_{j\le p} ||v_j||_2 = O_p(1)$, which is a uniformly bound. Thus,

$$B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})B_0 = O_p(1)\sum_{(i,j)\in J_L} |\Delta_{1,ij}| + O_p(1)\sum_{(i,j)\in J_U} |\Delta_{1,ij}|.$$

Now we consider the two terms $\sum_{(i,j)\in J_L} |\Delta_{1,ij}|$ and $\sum_{(i,j)\in J_U} |\Delta_{1,ij}|$ respectively. For $\sum_{(i,j)\in J_L} |\Delta_{1,ij}|$, according Lemma 3, we have

$$\frac{1}{2}\rho_{p,T}\sum_{(i,j)\in J_L} |\Delta_{1,ij}| \le pQ_1(\widehat{\Sigma}_e) - pQ_1(\Sigma_{e0}) - \Big[c_4\|\Delta\|_F^2 - 2\rho_{p,T}K_T$$
$$- \left(O_p\left(\sqrt{\frac{\log(p)}{T}}\right)\sqrt{p+D_p} + \rho_{p,T}\sqrt{D_p}\right)\|\Delta\|_F\Big]$$
$$= \mathcal{I}_1 + \mathcal{I}_2.$$

For $\mathcal{I}_1$, using (S4.5), we can obtain that $\mathcal{I}_1 \le O_p(\log(p) + p\sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)})$, and this result doesn't depend on the proof of this lemma. For $\mathcal{I}_2$, it is a univariate quadratic function on $||\Delta||_F$, which is bounded by $2\rho_{p,T}K_T + (O_p\left(\sqrt{\log(p)/T}\right)\sqrt{p+D_p} + \rho_{p,T}\sqrt{D_p})^2/(4c_4)$.

Thus we have:

$$\sum_{(i,j)\in J_L} |\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}| = O_p((\mathcal{I}_1 + \mathcal{I}_2)/\rho_{p,T})$$

$$= O_p((\log(p) + p\sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)})/\rho_{p,T} + K_T)$$

$$+ O_p((p + D_p)\log(p)/(T\rho_{p,T}) + \rho_{p,T}D_p).$$

Remind that under Assumptions 3 and 5, we have $\log(p)/p + \sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)}/p \ll \rho_{p,T} \ll p/D_p$ and $D_p \ll p\sqrt{T/\log(p)}$. Thus

$$D_p\log(p)/(T\rho_{p,T}) = (D_p\sqrt{\log(p)/T})(\sqrt{\log(p)/T}/\rho_{p,T}) = o(p)o(1) = o(p).$$

Combine with $K_T = o(p)$, we have

$$\sum_{(i,j)\in J_L} |\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}| = o_p(p).$$

Meanwhile, $\sum_{(i,j)\in J_U} |\Delta_{1,ij}|$ is bounded by

$$\sqrt{\sum_{i,j}(\Delta_{1,ij})^2 \sum_{i,j} 1^2_{\{(i,j)\in J_U\}}} = \sqrt{p + D_p}||\widehat{\Sigma}_e - \Sigma_{e0}||_F.$$

From (S4.6), we also see that

$$||\widehat{\Sigma}_e - \Sigma_{e0}||_F^2 = O_p(\rho_{p,T}K_T + \log(p) + p\sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)} + \rho_{p,T}^2 D_p).$$

Thus

$$(p + D_p)||\widehat{\Sigma}_e - \Sigma_{e0}||_F^2$$

$$= O_p((p + D_p)(\rho_{p,T}K_T + \log(p) + p\sqrt{\log(p)/T} + v_p^{-2(1+\varepsilon)} + \rho_{p,T}^2 D_p))$$

$$= O_p(p\rho_{p,T}K_T + p\log(p) + p^2\sqrt{\log(p)/T} + pv_p^{-2(1+\varepsilon)} + p\rho_{p,T}^2 D_p))$$

$$+ O_p(D_p\rho_{p,T}K_T + D_p\log(p) + D_p p\sqrt{\log(p)/T} + D_p v_p^{-2(1+\varepsilon)} + \rho_{p,T}^2 D_p^2)).$$

Now we bound $(p+D_p)||\widehat{\Sigma}_e-\Sigma_{e0}||_F^2$. The term $p\log(p)+p^2\sqrt{\log(p)/T}+pv_p^{-2(1+\varepsilon)}=o(p^2)$ is easy to verify. Assumption $\rho_{p,T}\ll p/K_T$ implies $p\rho_{p,T}K_T=o(p^2)$. Assumption $\rho_{p,T}\ll p/D_p$ and $K_T=o(p)$ imply $D_p\rho_{p,T}K_T=o(p^2)$ and $\rho_{p,T}^2D_p^2=o(p^2)$. Assumption of D_p implies that $D_p\log(p)+D_pp\sqrt{\log(p)/T}+D_pv_p^{-2(1+\varepsilon)}=o(p^2)$. Assumption $\rho_{p,T}\ll\sqrt{p/D_p}$ implies that $p\rho_{p,T}^2D_p=o(p^2)$. Thus $(p+D_p)||\widehat{\Sigma}_e-\Sigma_{e0}||_F^2=o_p(p^2)$ and derive the desired result

$$\sqrt{p+D_p}||\widehat{\Sigma}_e-\Sigma_{e0}||_F=o_p(p).$$

Thus we have

$$\sum_{(i,j)\in J_U}|\widehat{\Sigma}_{e,ij}-\Sigma_{e0,ij}|=o_p(p).$$

$$p^{-1}B_0^\top(\widehat{\Sigma}_e^{-1}-\Sigma_{e0}^{-1})B_0=p^{-1}(O_p(1)\sum_{(i,j)\in J_L}|\Delta_{1,ij}|+O_p(1)\sum_{(i,j)\in J_U}|\Delta_{1,ij}|)=o_p(1).$$

Lemma 5. *Under Assumptions 1,2 and 4,*

$$||(I_p-P_{J_p\widehat{B}})J_pB_0||_F=O_p(v_p^{-1}\max\{\sqrt{\frac{d_{p,T}}{\lambda}},1\}).$$

Proof: We prove Lemma 5 in two Steps. In Step I, we bound $||\widehat{\Theta}_A-\Theta_{A0}||_F$. Recall that $\Theta_A=\Theta_A(B,\Omega,\alpha)=J_pB\Omega B^\top J_p+\alpha\mathbf{1}_p^\top+\mathbf{1}_p\alpha^\top$, $\widehat{\Theta}_A=\Theta_A(\widehat{B},\widehat{\Omega},\widehat{\alpha})$ and $\Theta_{A0}=\Theta_A(B_0,\Omega_0,\alpha_0)$. In Step II, we use the relationship between $\widehat{\Theta}_A$ and $\widehat{B}$ to bound $||(I_p-P_{J_p\widehat{B}})J_pB_0||_F$.

Step I: By the definition of $(\widehat{B},\widehat{\Sigma}_e,\widehat{\Omega},\widehat{\alpha})$, we have $L(\widehat{B},\widehat{\Sigma}_e,\widehat{\Omega},\widehat{\alpha})\leq L(B_0,\widehat{\Sigma}_e,\Omega_0,\alpha_0)$.

That is

$$Q_1(\widehat{\Sigma}_e) + Q_2(\widehat{B}, \widehat{\Sigma}_e) + Q_3(\widehat{B}, \widehat{\Sigma}_e) + \frac{\lambda}{pT} L_A(\widehat{B}, \widehat{\Omega}, \widehat{\alpha}) \leq Q_1(\widehat{\Sigma}_e) + Q_2(B_0, \widehat{\Sigma}_e) + Q_3(B_0, \widehat{\Sigma}_e) + \frac{\lambda}{pT} L_A(B_0, \Omega_0, \alpha_0).$$

Using the definition of Q_2, we can easily derive that $Q_2(B_0, \widehat{\Sigma}_e) = 0$ and $Q_2(\widehat{B}, \widehat{\Sigma}_e) \geq 0$. Combining the results of Lemma 2, we have

$$L_A(\widehat{B}, \widehat{\Omega}, \widehat{\alpha}) - L_A(B_0, \Omega_0, \alpha_0) \leq O_p\left(\frac{\log(p)T}{\lambda} + \frac{p\sqrt{\log(p)T}}{\lambda}\right). \tag{S2.1}$$

Now we consider the term $L_A(\widehat{B}, \widehat{\Omega}, \widehat{\alpha}) - L_A(B_0, \Omega_0, \alpha_0)$. Let $A_1, P_1 \in \mathbb{R}^{p\times p}$ where $A_{1,ij} = A_{ij}1_{\{i<j\}}$ and $P_{1,ij} = P_{0,ij}1_{\{i<j\}}$, where we use P_0 to denote the true probability matrix of A. Notice that

$$\begin{aligned} L_A(B, \Omega, \alpha) &= -[tr(A_1\Theta_A) - \sum_{i<j} \log(1 + \exp(\Theta_{A,ij}))] \\ &= -[tr((A_1 - P_1)\Theta_A) + tr(P_1\Theta_A) - \sum_{i<j} \log(1 + \exp(\Theta_{A,ij}))] \\ &= -tr((A_1 - P_1)\Theta_A) - \sum_{i<j} [P_{0,ij}\Theta_{A,ij} - \log(1 + \exp(\Theta_{A,ij}))]. \end{aligned}$$

Using Taylor expansion, we have:

$$\begin{aligned} &\sum_{i<j} [P_{0,ij}\Theta_{A,ij} - \log(1 + \exp(\Theta_{A,ij}))] \\ &= -\sum_{i<j} \frac{\exp(\xi_{ij})}{(1 + \exp(\xi_{ij}))^2} (\Theta_{A,ij} - \Theta_{A0,ij})^2 + \sum_{i<j} [\Theta_{A0,ij}P_{0,ij} - \log(1 + \exp(\Theta_{A0,ij}))], \end{aligned}$$

where ξ_{ij} is between $\Theta_{A,ij}$ and $\Theta_{A0,ij}$. Thus

$$
\begin{aligned}
&L_A(\widehat{B},\widehat{\Omega},\widehat{\alpha}) - L_A(B_0,\Omega_0,\alpha_0)\\
&= -tr((A_1-P_1)(\widehat{\Theta}_A-\Theta_{A0})) + \sum_{i<j}\frac{\exp(\xi_{ij})}{(1+\exp(\xi_{ij}))^2}(\widehat{\Theta}_{A,ij}-\Theta_{A0,ij})^2\\
&\geq -tr((A_1-P_1)(\widehat{\Theta}_A-\Theta_{A0})) + \sum_{i<j}\min_{|\xi|\leq\max\{M_1,\delta_3\}}\frac{\exp(\xi)}{(1+\exp(\xi))^2}(\widehat{\Theta}_{A,ij}-\Theta_{A0,ij})^2\\
&= -tr((A_1-P_1)(\widehat{\Theta}_A-\Theta_{A0})) + \sum_{i<j}C_M(\widehat{\Theta}_{A,ij}-\Theta_{A0,ij})^2,
\end{aligned}
$$

where $C_M = \min_{|\xi|\leq\max\{M_1,\delta_3\}}\exp(\xi)/(1+\exp(\xi))^2$. Notice that

$$
|tr((A_1-P_1)(\widehat{\Theta}_A-\Theta_{A0}))| \leq ||A_1-P_1||_2\sqrt{\text{rank}(\widehat{\Theta}_A-\Theta_{A0})}||\widehat{\Theta}_A-\Theta_{A0}||_F,
$$

and $\sqrt{\text{rank}(\widehat{\Theta}_A-\Theta_{A0})} \leq \sqrt{2r+2}$. Using the results of Latała (2005), we have that $||A_1-P_1||_2 = O_p(\sqrt{p})$. Then

$$
L_A(\widehat{B},\widehat{\Omega},\widehat{\alpha}) - L_A(B_0,\Omega_0,\alpha_0) \geq \sum_{i<j}C_M(\widehat{\Theta}_{A,ij}-\Theta_{A0,ij})^2 - O_p(\sqrt{p})||\widehat{\Theta}_A-\Theta_{A0}||_F. \quad .
$$

Then we have:

$$
\begin{aligned}
&L_A(\widehat{B},\widehat{\Omega},\widehat{\alpha}) - L_A(B_0,\Omega_0,\alpha_0) + C_M\sum_{i\leq p}(\widehat{\Theta}_{A,ii}-\Theta_{A0,ii})^2/2 \\
&\geq C_M||\widehat{\Theta}_A-\Theta_{A0}||_F^2/2 - O_p(\sqrt{p})||\widehat{\Theta}_A-\Theta_{A0}||_F.
\end{aligned}
$$

Notice that $\sum_{i\leq p}(\widehat{\Theta}_{A,ii}-\Theta_{A0,ii})^2 \leq O_p(p)$ as $|\widehat{\Theta}_{A,ii}|$ and $|\Theta_{A0,ii}|$ are uniformly bounded, combine with the equation (S2.1), we have

$$
\frac{C_M}{2}||\widehat{\Theta}_A-\Theta_{A0}||_F^2 - O_p(\sqrt{p})||\widehat{\Theta}_A-\Theta_{A0}||_F \leq O_p(\frac{\log(p)T}{\lambda} + \frac{p\sqrt{\log(p)T}}{\lambda} + p).
$$

Based on quadratic function knowledge, we have

$$||\widehat{\Theta}_A - \Theta_{A0}||_F = O_p(\max(\sqrt{\frac{\log(p)T}{\lambda}}, \sqrt{p\frac{\sqrt{\log(p)T}}{\lambda}}, \sqrt{p})).$$

Step II: Let O_1 be the matrix such that $O_1^\top B_0^\top J_p B_0 O_1 = \Lambda_1$ is diagonal, and O_2 be the matrix such that $O_2^\top \widehat{B}^\top J_p \widehat{B} O_2 = \Lambda_2$ is diagonal.

According to Assumption 2 and using knowledge in OLS, we have

$$\begin{aligned}
&\frac{1}{\sqrt{Mp}}||(I_p - J_p\widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1}\widehat{B}^\top)J_pB_0||_F \\
&\leq \sigma_r(\Lambda_1^{-1/2})||(I_p - J_p\widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1}\widehat{B}^\top)J_pB_0||_F \\
&= \sigma_r(\Lambda_1^{-1/2}) \min_{X \in \mathbb{R}^{r\times r}} ||J_pB_0 - J_p\widehat{B}X||_F \\
&\leq \sigma_r(\Lambda_1^{-1/2})||J_pB_0 - J_p\widehat{B}O_2\Lambda_2^{-1/2}O_3\Lambda_1^{1/2}O_1^\top||_F \\
&\leq ||J_pB_0O_1\Lambda_1^{-1/2} - J_p\widehat{B}O_2\Lambda_2^{-1/2}O_3||_F,
\end{aligned} \tag{S2.2}$$

for any matrix $O_3 \in \mathbb{R}^{r\times r}$. According to Davis-Kahan Theorem (Davis and Kahan, 1970; Yu et al., 2015), there exist an orthogonal matrix O_3 and a constant M_2 such that

$$||J_pB_0O_1\Lambda_1^{-1/2} - J_p\widehat{B}O_2\Lambda_2^{-1/2}O_3||_F \leq \frac{M_2||J_p\widehat{B}\Omega\widehat{B}^\top J_p - J_pB_0\Omega B_0^\top J_p||_F}{\sigma_r(J_pB_0\Omega_0B_0^\top J_p) - \sigma_{r+1}(J_pB_0\Omega_0B_0^\top J_p)}.$$

Furthermore, as $J_p\mathbf{1}_p = \mathbf{0}_p$, we have:

$$
\begin{aligned}
||\widehat{\Theta}_A - \Theta_{A0}||_F^2 &= ||J_p\widehat{B}\widehat{\Omega}\widehat{B}^\top J_p - J_pB_0\Omega_0B_0^\top J_p||_F^2 + ||\mathbf{1}_p(\widehat{\alpha}-\alpha_0)^\top + (\widehat{\alpha}-\alpha_0)\mathbf{1}_p^\top||_F^2 \\
&+ 2tr((J_p\widehat{B}\widehat{\Omega}\widehat{B}^\top J_p - J_pB_0\Omega_0B_0^\top J_p)(\mathbf{1}_p(\widehat{\alpha}-\alpha_0)^\top + (\widehat{\alpha}-\alpha_0)\mathbf{1}_p^\top)) \\
&= ||J_p\widehat{B}\widehat{\Omega}\widehat{B}^\top J_p - J_pB_0\Omega_0B_0^\top J_p||_F^2 + ||\mathbf{1}_p(\widehat{\alpha}-\alpha_0)^\top + (\widehat{\alpha}-\alpha_0)\mathbf{1}_p^\top||_F^2 \\
&+ 2tr((\widehat{B}\widehat{\Omega}\widehat{B}^\top - B_0\Omega_0B_0^\top)J_p(\mathbf{1}_p(\widehat{\alpha}-\alpha_0)^\top + (\widehat{\alpha}-\alpha_0)\mathbf{1}_p^\top)J_p) \\
&= ||J_p\widehat{B}\widehat{\Omega}\widehat{B}^\top J_p - J_pB_0\Omega_0B_0^\top J_p||_F^2 + ||\mathbf{1}_p(\widehat{\alpha}-\alpha_0)^\top + (\widehat{\alpha}-\alpha_0)\mathbf{1}_p^\top||_F^2 \\
&\geq ||J_p\widehat{B}\widehat{\Omega}\widehat{B}^\top J_p - J_pB_0\Omega_0B_0^\top J_p||_F^2,
\end{aligned}
$$

which implies that $||J_p\widehat{B}\Omega\widehat{B}^\top J_p - J_pB_0\Omega B_0^\top J_p||_F$ is bounded by $||\widehat{\Theta}_A - \Theta_{A0}||_F$. By Assumption 3, $\sigma_{r+1}(J_pB_0\Omega_0B_0^\top J_p) = 0$ and $\sigma_r(J_pB_0\Omega_0B_0^\top J_p) \geq \sigma_r^2(J_pB_0)\sigma_r(\Omega_0) \geq c_3mpv_p$. Thus,

$$
||J_pB_0O_1\Lambda_1^{-1/2} - J_p\widehat{B}O_2\Lambda_2^{-1/2}O_3||_F = O_p(\frac{1}{pv_p}\max(\sqrt{\frac{\log(p)T}{\lambda}}, \sqrt{p\frac{\sqrt{\log(p)T}}{\lambda}}, \sqrt{p})).
$$

By (S2.2), we have:

$$
\begin{aligned}
||(I_p - J_p\widehat{B}(\widehat{B}^\top J_p\widehat{B})^{-1}\widehat{B}^\top J_p)J_pB_0||_F &= O_p(v_p^{-1}\max\{\sqrt{\frac{\log(p)T}{\lambda p}}, \sqrt{\frac{\sqrt{\log(p)T}}{\lambda}}, 1\}) \\
&= O_p(v_p^{-1}\max\{\sqrt{\frac{d_{p,T}}{\lambda}}, 1\}).
\end{aligned}
$$

Lemma 6. *Under Assumption 3 and without assuming the lower bound of $\sigma_r(\Omega)$, we have*

$$
\sup_{(B,\Sigma_e,\Omega,\alpha)\in\Xi_\delta} (L_A(B,\Omega,\alpha) - L_A(B_0,\Omega_0,\alpha_0))^- = O_p(p).
$$

Proof: Using result in Lemma 5, for any $(B, \Sigma_e, \Omega, \alpha) \in \Xi_\delta$, we have:

$$L_A(B, \Omega, \alpha) - L_A(B_0, \Omega_0, \alpha_0) \geq \frac{C_M}{2}||\Theta_A - \Theta_{A0}||_F^2 + O_p(p) \\ - ||A_1 - P_1||_2\sqrt{\text{rank}(\Theta_A - \Theta_{A0})}||\Theta_A - \Theta_{A0}||_F.$$

Then we have

$$\sup_{(B,\Sigma_e,\Omega,\alpha)\in\Xi_\delta}(L_A(B, \Omega, \alpha) - L_A(B_0, \Omega_0, \alpha_0))^- \\ \leq (||A_1 - P_1||_2\sqrt{\text{rank}(\Theta_A - \Theta_{A0})})^2/(2C_M) + O_p(p) = O_p(p).$$

Lemma 7. *For any random function $Z_n(x, \omega) : R \times \Omega \to R$ satisfies $Z_n(x_n, \omega) \to_p 0$ for any $x_n \ll a_n$ or $x_n \gg b_n$, where $0 < b_n \ll a_n$. Then for any nonrandom sequence $\{\lambda_n\}_{n=1}^\infty$, we have $Z_n(\lambda_n, \omega) \to_p 0$.*

Proof:

For any λ_n and ε, we have

$$\Pr(|Z_n(\lambda_n, \omega)| > \varepsilon) \\ \leq \Pr(\max(|Z_n(\min(\lambda_n, \sqrt{a_n b_n}), \omega)|, |Z_n(\max(\lambda_n, \sqrt{a_n b_n}), \omega)|) > \varepsilon) \\ \leq \Pr(|Z_n(\min(\lambda_n, \sqrt{a_n b_n}), \omega)| > \varepsilon) + \Pr(|Z_n(\max(\lambda_n, \sqrt{a_n b_n}), \omega)| > \varepsilon) \\ \to 0,$$

as $\min(\lambda_n, \sqrt{a_n b_n}) \ll a_n$ and $\max(\lambda_n, \sqrt{a_n b_n}) \gg b_n$. Thus $Z_n(\lambda_n, \omega) \to_p 0$.

Lemma 8. *Assume matrix $X \in \mathbb{R}^{k\times k}$ satisfy $X^\top X - I_r = 0$ and $X\Phi_1 X^\top = \Phi_2$ for diagonal matrices Φ_1 and Φ_2, where Φ_2 has distinct diagonal elements. Then X is diagonal with $X_{ii} = 1$ or -1.*

Lemma 9. *For any matrix $Z_1, Z_2 \in \mathbb{R}^{p\times r}$, we have:*

$$\min_{OO^\top = O^\top O = I_r} ||Z_1 - Z_2 O||_F^2 \leq (2\sqrt{2}-2)^{-1}\sigma_r^{-1}(Z_2 Z_2^\top)||Z_1 Z_1^\top - Z_2 Z_2^\top||_F^2.$$

Lemma 10. *Assume $z_t = g(f_t, e_t)$ for measurable function g and there exists a positive C_z such that $E(z_t^6|F)^{1/6} \leq C_z$ where C_z doesn't depend on p, t, T and F, $E(z_t|F) = 0$. Under Assumptions 1, 6 and 7, we have $E[(T^{-1/2}\sum_{t=1}^T z_t)^4|F]$ is bounded by a constant that only depend on C_z, a_3', d_3'.*

Proof: We denote $E_{\cdot|F}(X) = E(X|F)$ for random variable X. Notice that

$$\begin{aligned} E_{\cdot|F}(T^{-1/2}\sum_{t=1}^T z_t)^4 &= T^{-2}\sum_{i,j,k,l} E_{\cdot|F}(z_i z_j z_k z_l) \\ &= T^{-2}\sum_{1\leq t_1\leq t_2\leq T}\ \sum_{\min\{i,j,k,l\}=t_1,\max\{i,j,k,l\}=t_2} E_{\cdot|F}(z_i z_j z_k z_l). \end{aligned}$$

Now we consider the term $\sum_{\min\{i,j,k,l\}=t_1,\max\{i,j,k,l\}=t_2} |E_{\cdot|F}(z_i z_j z_k z_l)|$. We denote $g_1(x_1, x_2, x_3, x_4) = \max\{|x_{(2)} - x_{(1)}|, |x_{(4)} - x_{(3)}|\}$ where $x_{(i)}$ is the ith smallest element in $\{x_1, x_2, x_3, x_4\}$. We have

$$\begin{aligned} &\sum_{\min\{i,j,k,l\}=t_1,\max\{i,j,k,l\}=t_2} |E_{\cdot|F}(z_i z_j z_k z_l)| \\ &= \sum_{d=0}^{\infty}\ \sum_{\min\{i,j,k,l\}=t_1,\max\{i,j,k,l\}=t_2, g_1(i,j,k,l)=d} |E_{\cdot|F}(z_i z_j z_k z_l)|. \end{aligned}$$

For each d, we first bound the $|E_{\cdot|F}(z_i z_j z_k z_l)|$. Without loss of generality, we assume $i \leq j \leq k \leq l$ and $|i-j| = d$. Notice $\{f_t\}$ is independent to

$\{e_t\}$, we have:

$$
\begin{aligned}
&|E(z_i z_j z_k z_l|F=\{f_t\})| \\
&= |[E \prod_{t=i,j,k,l} g(x_t,e_t)]\Big|_{\{x_t=f_t, t=i,j,k,l\}}| \\
&= |[cov(g(x_i,e_i), \prod_{t=j,k,l} g(x_t,e_t))]\Big|_{\{x_t=f_t, t=i,j,k,l\}}| \\
&\le \rho(d)([var(g(x_i,e_i))]^{1/2}[var(g(x_j,e_j)g(x_k,e_k)g(x_l,e_l))]^{1/2})\Big|_{\{x_t=f_t, t=i,j,k,l\}} \\
&\le \rho(d)([Eg^2(x_i,e_i)]^{1/2}[Eg^2(x_j,e_j)g^2(x_k,e_k)g^2(x_l,e_l)]^{1/2})\Big|_{\{x_t=f_t, t=i,j,k,l\}} \\
&\le \rho(d)[Eg^2(x_i,e_i)]^{1/2}[Eg^6(x_j,e_j)]^{1/6}[Eg^6(x_k,e_k)]^{1/6}[Eg^6(x_l,e_l)]^{1/6}\Big|_{\{x_t=f_t, t=i,j,k,l\}} \\
&\le \rho(d)([E_{\cdot|F} g^2(f_i,e_i)]^{1/2}[E_{\cdot|F} g^6(f_j,e_j)]^{1/6} E_{\cdot|F}[g^6(f_k,e_k)]^{1/6} E_{\cdot|F}[g^6(f_l,e_l)]^{1/6} \\
&\le C_z^4 \exp(-a_3' d^{r_3'}).
\end{aligned}
$$

Meanwhile we should notice that $\#\{(i,j,k,l) : \min\{i,j,k,l\} = k_1, \max\{i,j,k,l\} = k_2, g_1(i,j,k,l) = d\} \le 24(2d+2)^2$. Thus we have

$$
\begin{aligned}
&|\sum_{d=0}^{\infty} \sum_{\min\{i,j,k,l\}=k_1, \max\{i,j,k,l\}=k_2, g_1(i,j,k,l)=d} E_{\cdot|F}(z_i z_j z_k z_l)| \\
&\le \sum_{d=0}^{\infty} 24(2d+2)^2 C_z^4 \exp(-a_3' d^{r_3'}).
\end{aligned}
$$

This is bounded by a constant only depend on C_z, a_3', r_3'. Thus

$$
|\sum_{1\le k_1,k_2 \le T} \sum_{d=0}^{\infty} \sum_{\min\{i,j,k,l\}=k_1, \max\{i,j,k,l\}=k_2, g_1(i,j,k,l)=d} E_{\cdot|F}(z_i z_j z_k z_l)|,
$$

is bounded by T^2 multiplied by a constant, which implies $E_{\cdot|F}(T^{-1/2}\sum_{t=1}^{T} z_t)^4$ is uniformly bounded by a constant.

S.3 Proof of Theorem 1

Since $J_pB\Omega B^\top J_p + \mathbf{1}_p\alpha^\top + \alpha\mathbf{1}_p^\top = J_pB_\star\Omega_\star B_\star^\top J_p + \mathbf{1}_p\alpha_\star^\top + \alpha_\star\mathbf{1}_p^\top$, we have

$$J_pB\Omega B^\top J_p - J_pB_\star\Omega_\star B_\star^\top J_p = \mathbf{1}_p(\alpha_\star - \alpha)^\top + (\alpha_\star - \alpha)\mathbf{1}_p^\top,$$

and hence

$$J_p(J_pB\Omega B^\top J_p - J_pB_\star\Omega_\star B_\star^\top J_p)J_p = J_p(\mathbf{1}_p(\alpha_\star - \alpha)^\top + (\alpha_\star - \alpha)\mathbf{1}_p^\top)J_p.$$

Notice that $J_p^2 = J_p$ and $J_p\mathbf{1}_p = 0$, we have

$$J_pB\Omega B^\top J_p - J_pB_\star\Omega_\star B_\star^\top J_p = \mathbf{0}_{p\times p},$$

and

$$\mathbf{1}_p(\alpha_\star - \alpha)^\top + (\alpha_\star - \alpha)\mathbf{1}_p^\top = J_pB\Omega B^\top J_p - J_pB_\star\Omega_\star B_\star^\top J_p = \mathbf{0}_{p\times p}.$$

Notice that $\mathrm{diag}(\mathbf{1}_p(\alpha_\star - \alpha)^\top + (\alpha_\star - \alpha)\mathbf{1}_p^\top) = 2(\alpha_\star - \alpha) = \mathbf{0}_p$, we have $\alpha_\star = \alpha$.

Now we back to $J_pB\Omega B^\top J_p - J_pB_\star\Omega_\star B_\star^\top J_p = \mathbf{0}_{p\times p}$. It is easy to verify that $\sigma_r(J_pB\Omega B^\top J_p) \geq \sigma_r^2(J_pB)\sigma_r(\Omega) \neq 0$. Thus

$$\mathrm{rank}(J_pB) = \mathrm{rank}(J_pB\Omega B^\top J_p) = r.$$

We will use this result later.

Denote $\mathrm{span}(X) = \{Xa : a \text{ is a vector}\}$. Then $\mathrm{span}(J_pB\Omega B^\top J_p) \subset \mathrm{span}(J_pB)$. By the fact $\mathrm{rank}(J_pB) = \mathrm{rank}(J_pB\Omega B^\top J_p)$, we have $\mathrm{span}(J_pB\Omega B^\top J_p) =$

span(J_pB). Similarly we have span(J_pB) = span($J_pB\Omega B^\top J_p$) = span($J_pB_\star\Omega_\star B_\star^\top J_p$) = span($J_pB_\star$). Thus there exists a matrix X_0 such that $J_pBX_0 = J_pB_\star$.

Now we prove that there exists an orthogonal matrix $O_\star$ such that $B = B_\star O_\star$. Consider that $B = (J_pB, \mathbf{1}_p)(I_r, x_0)^\top = (J_pB, \mathbf{1}_p)W$ and $B_\star = (J_pB_\star, \mathbf{1}_p)(I_r, x_\star)^\top = (J_pB, \mathbf{1}_p)(X_0^\top, x_\star)^\top = (J_pB, \mathbf{1}_p)W_\star$. We only need to prove there exists an orthogonal matrix $O_\star$ such that $W = W_\star O_\star$. From lemma 9, we see that we only need to prove $WW^\top = W_\star W_\star^\top$.

Denote $G_0 = (J_pB, \mathbf{1}_p)$. For matrix $X \in \mathbb{R}^{p_1\times p_2}$, we denote $[X]_{i\cdot} = (X_{i1}, ..., X_{ip_2})$. We should notice that

$$\|J_pB\Omega[J_pB]_{i\cdot}^\top\|_F = (\sum_{j=1}^{p}([J_p\Theta_A J_p]_{ij})^2)^{1/2} \le 4(p\sup_{i,j}(\Theta_{A,ij})^2)^{1/2} \le 4\delta_3 p^{1/2},$$

and

$$\|J_pB\Omega[J_pB]_{i\cdot}^\top\|_2 \ge \sigma_r(J_pB)\sigma_r(\Omega)\|[J_pB]_{i\cdot}\|_2 \ge \delta_2^{-1/2}\delta_5 p^{1/2}v_p\|[J_pB]_{i\cdot}\|_2.$$

Thus

$$\sup_i \|[J_pB]_{i\cdot}\|_2 \le 4\delta_3\delta_2^{1/2}\delta_5^{-1}v_p^{-1} := K_1(\delta)v_p^{-1}.$$

Now we use the results presented above to prove that we can extract about $O(pv_p^2)$ of $(r+1)\times(r+1)$ full rank submatrices from G_0. First, we have

$$||G_{0,i\cdot}||_2 \le ||[J_pB]_{i\cdot}||_2 + 1 \le (K_1(\delta) + \sup_{p\in\mathbb{N}^+} v_p)v_p^{-1} \le (K_1(\delta) + K_2)v_p^{-1},$$

where $K_2 = \sup_{p\in\mathbb{N}^+} v_p = O(1)$. Meanwhile, notice that

$$\sigma_{r+1}(G_0) = \sigma_{r+1}^{1/2}(G_0^\top G_0) = \sigma_{r+1}^{1/2}(\begin{pmatrix} B^\top J_p B & \mathbf{0}_{p\times 1} \\ \mathbf{0}_{1\times p} & p \end{pmatrix}) \geq \sqrt{\min(\delta_2, 1)p}.$$

Denote $G^{(0)} = G_0$, $\mathbb{I}^{(0)} = \emptyset$. We consider an iterative procedure: For the k-th step, we draw $\mathbb{I}^{(k)} = \{i_1^{(k)}, ..., i_{r+1}^{(k)}\} \subset \{1, 2, ..., p\}$ such that $\mathbb{I}^{(k)} \cap \bigcup_{i=0}^{k-1} \mathbb{I}^{(i)} = \emptyset$ and $G^{(k)} = ([G^{(0)}]_{i_1^{(k)}\cdot}^\top, ..., [G^{(0)}]_{i_{r+1}^{(k)}\cdot}^\top)$ is full rank. We denote $G^{(-k)}$ is the matrix after removing $G^{(1)}, ..., G^{(k)}$ from $G^{(0)}$. We denote $\mathbb{I}_k = \bigcup_{i=0}^{k} \mathbb{I}^{(i)}$. Notice that after the k-th step, we have

$$\begin{aligned}
\sigma_{r+1}^2(G^{(-k)}) &= \sigma_{r+1}^2(G^{(-k)\top} G^{(-k)}) \\
&= \lambda_{r+1}(\sum_{i\in\{1,2,...,p\}-\mathbb{I}_k} [G^{(0)}]_{i\cdot}^\top [G^{(0)}]_{i\cdot}) \\
&= \lambda_{r+1}(\sum_{i=1}^{p} [G^{(0)}]_{i\cdot}^\top [G^{(0)}]_{i\cdot} - \sum_{i\in\mathbb{I}_k} [G^{(0)}]_{i\cdot}^\top [G^{(0)}]_{i\cdot}) \\
&\geq \lambda_{r+1}(\sum_{i=1}^{p} [G^{(0)}]_{i\cdot}^\top [G^{(0)}]_{i\cdot}) - \lambda_1(\sum_{i\in\mathbb{I}_k} [G^{(0)}]_{i\cdot}^\top [G^{(0)}]_{i\cdot}) \\
&\geq \lambda_{r+1}(G^{(0)\top} G^{(0)}) - \sum_{i\in\mathbb{I}_k} ||[G^{(0)}]_{i\cdot}||_2^2 \\
&\geq \min(\delta_2^{-1}, 1)p - (r+1)k(K_1(\delta) + K_2)^2 v_p^{-2},
\end{aligned}$$

which implies we can repeat this procedure until $k = \lfloor \min(\delta_2^{-1}, 1)(r+1)^{-1}(K_1(\delta) + K_2)^{-2} p v_p^2 \rfloor > K_3(\delta) p v_p^2$, where $K_3(\delta) = 2^{-1}\min(\delta_2^{-1}, 1)(r+1)^{-1}(K_1(\delta) + K_2)^{-2}$ for some large p. If $WW^\top \neq W_\star W_\star^\top$, for all pairs of $0 < k_1, k_2 \leq \lfloor K_3(\delta) p v_p^2 \rfloor + 1, k_1 \neq k_2$, we have $G^{(k_1)}(WW^\top - W_\star W_\star^\top) G^{(k_2)\top} \neq$

$\mathbf{0}_{(r+1)\times(r+1)}$, which implies there exists $(i_{k_1}, i_{k_2}) \in \{i_1^{(k_1)}, ..., i_{r+1}^{(k_1)}\} \times \{i_1^{(k_2)}, ..., i_{r+1}^{(k_2)}\}$ such that $(BB^\top - B_\star B_\star^\top)_{i_{k_1} i_{k_2}} \neq 0$, and then

$$\sum_{i \neq j} 1_{\{(BB^\top - B_\star B_\star^\top)_{ij} \neq 0\}} \geq (\max(K_3(\delta) p v_p^2 - 1, 0))^2.$$

Denote $\mathrm{Ndiag}(X) \in \mathbb{R}^{p\times p}$, where $[\mathrm{Ndiag}(X)]_{ij} = X_{ij} 1_{\{i\neq j\}}$ for $X \in \mathbb{R}^{p\times p}$. Notice that $\|\Sigma_e\|_0 - p = \|\mathrm{Ndiag}(\Sigma_e)\|_0$ and $\|\Sigma_{e\star}\|_0 - p = \|\mathrm{Ndiag}(\Sigma_{e\star})\|_0$, we choose $C(\delta) = 4^{-1}(K_3(\delta) - \sup_{p\geq p_0}(p^{-1} v_p^{-2}))^2$ for large enough p_0 such that $K_3(\delta) - \sup_{p\geq p_0}(p^{-1} v_p^{-2}) > 0$. Then for $p \geq p_0$, we have

$$\begin{aligned} \max\{\|\Sigma_e\|_0, \|\Sigma_{e\star}\|_0\} - p &\geq 2^{-1}\|\mathrm{Ndiag}(\Sigma_e - \Sigma_{e\star})\|_0 \\ &= 2^{-1}\|\mathrm{Ndiag}(BB^\top - B_\star B_\star^\top)\|_0 \\ &\geq 2^{-1}(K_3(\delta) p v_p^2 - 1)^2 \geq 2C(\delta) p^2 v_p^4. \end{aligned}$$

This contradict to $\max\{\|\Sigma_e\|_0, \|\Sigma_{e\star}\|_0\} - p \leq C(\delta) p^2 v_p^4$. Thus we have $WW^\top = W_\star W_\star^\top$ and $BB^\top - B_\star B_\star^\top = G_0(WW^\top - W_\star W_\star^\top)G_0^\top = \mathbf{0}_{p\times p}$. Thus there exists an orthogonal matrix $O_\star \in \mathbb{R}^{r\times r}$ such that $||B - B_\star O_\star||_F = 0$ and $\Sigma_e - \Sigma_{e\star} = -(BB^\top - B_\star B_\star^\top) = \mathbf{0}_{p\times p}$.

Finally, we consider Ω. Remind that $J_p B \Omega B^\top J_p - J_p B_\star \Omega_\star B_\star^\top J_p = \mathbf{0}_{p\times p}$. We have $J_p B(\Omega - O_\star \Omega_\star O_\star^\top) B^\top J_p = \mathbf{0}_{p\times p}$. As $\mathrm{rank}(J_p B) = r$, we have $J_p B(\Omega - O_\star \Omega_\star O_\star^\top) B^\top J_p = \mathbf{0}_{p\times p}$ if and only if $\Omega - O_\star \Omega_\star O_\star^\top = \mathbf{0}_{r\times r}$.

S.4 Proof of Theorem 2

We divide the proof into two Parts. In Part I, we prove the consistency result when $\lambda \gg \min\{v_p^{2\varepsilon}, \log^{-\varepsilon/(1+\varepsilon)}(p)\}d_{p,T}$. In Part II, we illustrate the consistency result when $\lambda \ll d_{p,T}$. Finally, as $\min\{v_p^{2\varepsilon}, \log^{-\varepsilon/(1+\varepsilon)}(p)\}d_{p,T} \ll d_{p,T}$, we directly use Lemma 7 for the desired result.

Part I: In this Part, we use three Steps to prove the desired result. In Step I, we prove the consistency result of $\widehat{\Sigma}_e$. In Step II, we bounded an important term $R = (\widehat{B} - B_0)^\top \widehat{\Sigma}_e^{-1}\widehat{B}(\widehat{B}^\top \widehat{\Sigma}_e^{-1}\widehat{B})^{-1}$ which has been used in (Bai and Liao, 2016). In Step III, we prove the consistency result.

Step I: The main idea in this Step is to bound the $pQ_1(\widehat{\Sigma}_e) - pQ_1(\Sigma_{e0})$ and apply Lemma 3 to derive the consistency result of $\widehat{\Sigma}_e$.

Consider that for any $\gamma \in \mathbb{R}^r$ such that $\sigma_1(J_p\widehat{B} + \mathbf{1}_p\gamma^\top) \le \sqrt{\delta_1 p}$ and $\sigma_r(J_p\widehat{B} + \mathbf{1}_p\gamma^\top) \ge \sqrt{\delta_1^{-1}p}$, which implies that $(J_p\widehat{B} + \mathbf{1}_p\gamma^\top, \Sigma_{e0}, \widehat{\Omega}, \widehat{\alpha}) \in \Xi_\delta$. We have

$$L(\widehat{B}, \widehat{\Sigma}_e, \widehat{\Omega}, \widehat{\alpha}) \le L(J_p\widehat{B} + \mathbf{1}_p\gamma^\top, \Sigma_{e0}, \widehat{\Omega}, \widehat{\alpha}).$$

That is

$$\begin{aligned}
&pQ_1(\widehat{\Sigma}_e) + pQ_2(\widehat{B}, \widehat{\Sigma}_e) + pQ_3(\widehat{B}, \widehat{\Sigma}_e) + \lambda T^{-1}L_A(\widehat{B}, \widehat{\Omega}, \widehat{\alpha}) \\
&\le pQ_1(\Sigma_{e0}) + pQ_2(J_p\widehat{B} + \mathbf{1}_p\gamma^\top, \Sigma_{e0}) + pQ_3(J_p\widehat{B} \\
&+ \mathbf{1}_p\gamma^\top, \Sigma_{e0}) + \lambda T^{-1}L_A(J_p\widehat{B} + \mathbf{1}_p\gamma^\top, \widehat{\Omega}, \widehat{\alpha}).
\end{aligned}$$

It is easy to verify that

$$L_A(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\widehat{\Omega},\widehat{\alpha})=L_A(\widehat{B},\widehat{\Omega},\widehat{\alpha}).$$

Combined with Lemma 2 that $\sup_{(B,\Sigma_e,\Omega,\alpha)\in\Xi_\delta}|Q_3(B,\Sigma_e)|=O_p(\log(p)/p+\sqrt{\log(p)/T})$, we have

$$\begin{aligned}
&(pQ_1(\widehat{\Sigma}_e)-pQ_1(\Sigma_{e0}))+(pQ_2(\widehat{B},\widehat{\Sigma}_e)-pQ_2(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\Sigma_{e0}))\\
&\le pQ_3(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\Sigma_{e0})-pQ_3(\widehat{B},\widehat{\Sigma}_e)+\lambda T^{-1}(L_A(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\widehat{\Omega},\widehat{\alpha})\\
&\quad -L_A(\widehat{B},\widehat{\Omega},\widehat{\alpha}))\\
&\le O_p(\log(p)+p\sqrt{\frac{\log(p)}{T}}).
\end{aligned} \tag{S4.3}$$

First, we bound the $pQ_2(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\Sigma_{e0})$. Denote $\gamma_0=B_0^\top\mathbf{1}_p/p$ and $B^*=J_p\widehat{B}+\mathbf{1}_p\gamma^\top$.

$$\begin{aligned}
&pQ_2(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\Sigma_{e0})\\
&=||(I_p-\Sigma_{e0}^{-1/2}B^*(B^{*T}\Sigma_{e0}^{-1}B^*)^{-1}B^{*T}\Sigma_{e0}^{-1/2})\Sigma_{e0}^{-1/2}B_0||_F^2\\
&=\min_{X\in\mathbb{R}^{r\times r}}||\Sigma_{e0}^{-1/2}B_0-\Sigma_{e0}^{-1/2}B^*X||_F^2\\
&\le||\Sigma_{e0}^{-1/2}B_0-\Sigma_{e0}^{-1/2}B^*(\widehat{B}^\top J_p\widehat{B})^{-1}\widehat{B}^\top J_pB_0||_F^2\\
&\le\sigma_1(\Sigma_{e0}^{-1})||J_pB_0-J_p\widehat{B}(\widehat{B}^\top J_p\widehat{B})^{-1}\widehat{B}^\top J_pB_0+\mathbf{1}_p(\gamma_0^\top-\gamma^\top(\widehat{B}^\top J_p\widehat{B})^{-1}\widehat{B}^\top J_pB_0)||_F^2\\
&\le 2c^{-1}||J_pB_0-J_p\widehat{B}(\widehat{B}^\top J_p\widehat{B})^{-1}\widehat{B}^\top J_pB_0||_F^2\\
&+2c^{-1}||\mathbf{1}_p(\gamma_0^\top-\gamma^\top(\widehat{B}^\top J_p\widehat{B})^{-1}\widehat{B}^\top J_pB_0)||_F^2.
\end{aligned}$$

Set $\gamma=(B_0^\top J_p\widehat{B}(\widehat{B}^\top J_p\widehat{B})^{-1})^{-1}\gamma_0$ such that $||\mathbf{1}_p(\gamma_0^\top-\gamma^\top(\widehat{B}^\top J_p\widehat{B})^{-1}\widehat{B}^\top J_pB_0)||_F^2=$

0. To choose such a γ, we need to illustrate that it is well defined. First, we need to prove that for sufficiently large p and T, $B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1}$ is invertible. Notice that

$$B_0^\top J_p B_0 = B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1} \widehat{B}^\top J_p B_0 + (B_0^\top J_p B_0 - B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1} \widehat{B}^\top J_p B_0).$$

Now we bound $B_0^\top J_p B_0 - B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1} \widehat{B}^\top J_p B_0 = B_0^\top J_p (I_p - P_{J_p \widehat{B}}) J_p B_0$. First, we need to bound $||(I_p - P_{J_p\widehat{B}}) J_p B_0||_F^2$. Using the result in Lemma 5, we have

$$\begin{aligned} ||(I_p - P_{J_p\widehat{B}}) J_p B_0||_F^2 &= (O_p(v_p^{-1} \max\{\sqrt{\frac{d_{p,T}}{\lambda}}, 1\})^2 \\ &\leq O_p(\max\{v_p^{-1} \max\{v_p^{-\varepsilon}, \log^{\varepsilon/2(1+\varepsilon)}(p))\}, v_p^{-1}\}))^2 \\ &= O_p((\max\{v_p^{-(1+\varepsilon)}, \log^{1/2}(p)\})^2) \\ &= O_p(v_p^{-2(1+\varepsilon)} + \log(p)) = o_p(p). \end{aligned} \tag{S4.4}$$

Thus, we have the term $B_0^\top J_p B_0 - B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1} \widehat{B}^\top J_p B_0 = B_0^\top J_p (I_p - P_{J_p\widehat{B}}) J_p B_0$ is semi-positive definite and its max eigenvalue is bounded by $tr(B_0^\top J_p B_0 - B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1} \widehat{B^\top} J_p B_0) = ||(I_p - P_{J_p\widehat{B}}) J_p B_0||_F^2 = o_p(p)$ by the result in (S4.4). Thus

$$\begin{aligned} \sigma_r^2(B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1}) \sigma_1(\widehat{B}^\top J_p \widehat{B}) &\geq \sigma_r(B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1} \widehat{B}^\top J_p B_0) \\ &\geq \sigma_r(B_0^\top J_p B_0) - o_p(p) \geq mp - o_p(p). \end{aligned}$$

Thus $\sigma_r(B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1}) \geq \sqrt{m/\delta_2} + o_p(1)$, which implies the matrix $B_0^\top J_p \widehat{B}(\widehat{B}^\top J_p \widehat{B})^{-1}$ is invertible. Then, we need to test the upper and lower bound of $\sigma_1(J_p\widehat{B}+\mathbf{1}_p\gamma^\top)$ and $\sigma_r(J_p\widehat{B}+\mathbf{1}_p\gamma^\top)$. We have $\sigma_1(J_p\widehat{B}+\mathbf{1}_p\gamma^\top) \leq ||J_p\widehat{B}||_2 + ||\mathbf{1}_p||_2||\gamma_0||_2||(B_0^\top J_p\widehat{B}(\widehat{B}^\top J_p\widehat{B})^{-1})^{-1}||_2 \leq \sqrt{\delta_2 p} + \sqrt{C_1 rp\delta_2/m} + \sqrt{\varepsilon p}$ (ε can be replaced by any small positive number) by noticing that $p||\gamma_0||_F^2 = ||\mathbf{1}_p\gamma_0^\top||_F^2 \leq tr(J_pB_0B_0^\top J_p) + tr(\mathbf{1}_p\gamma_0^\top\gamma_0\mathbf{1}_p^\top) = ||J_pB_0 + \mathbf{1}_p\gamma_0^\top||_F^2 \leq C_1rp$. We assume δ_1 is large enough. Thus $\sigma_1(J_p\widehat{B} + \mathbf{1}_p\gamma^\top) \leq \sqrt{\delta_1 p}$. For lower bound, we have

$$\begin{aligned}
\sigma_r(J_p\widehat{B} + \mathbf{1}_p\gamma^\top) &= \sqrt{\sigma_r((J_p\widehat{B} + \mathbf{1}_p\gamma^\top)^\top(J_p\widehat{B} + \mathbf{1}_p\gamma^\top))} \\
&= \sqrt{\sigma_r(\widehat{B}^\top J_p\widehat{B} + p\gamma\gamma^\top)} \\
&\geq \sqrt{\sigma_r(\widehat{B}^\top J_p\widehat{B})} \geq \sqrt{\delta_2^{-1}p} \geq \sqrt{\delta_1^{-1}p}.
\end{aligned}$$

Thus we prove that γ is well defined. For the chosen γ, we have

$$pQ_2(J_p\widehat{B} + \mathbf{1}_p\gamma^\top, \Sigma_{e0}) \leq 2c^{-1}||(I_p - P_{J_p\widehat{B}})J_pB_0||_F^2 = O_p(v_p^{-2(1+\varepsilon)} + \log(p)).$$

Combined with (S4.3) and notice that $pQ_2(\widehat{B}, \widehat{\Sigma}_e) \geq 0$, we have

$$pQ_1(\widehat{\Sigma}_e) - pQ_1(\Sigma_{e0}) \leq O_p(\log(p) + p\sqrt{\frac{\log(p)}{T}} + v_p^{-2(1+\varepsilon)}). \tag{S4.5}$$

Using Lemma 3, we have

$$
\begin{aligned}
& c_4\|\Delta\|_F^2 - 2\rho_{p,T}K_T \\
& - \left(O_p\left(\sqrt{\frac{\log(p)}{T}}\right)\sqrt{p+D_p} + \rho_{p,T}\sqrt{D_p}\right)\|\Delta\|_F \\
& \leq \frac{1}{2}\rho_{p,T}\sum_{(i,j)\in J_L}\left|\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}\right| + c_4\|\Delta\|_F^2 - 2\rho_{p,T}K_T \\
& - \left(O_p\left(\sqrt{\frac{\log(p)}{T}}\right)\sqrt{p+D_p} + \rho_{p,T}\sqrt{D_p}\right)\|\Delta\|_F \\
& \leq O_p(\log(p) + p\sqrt{\frac{\log(p)}{T}} + v_p^{-2(1+\varepsilon)}).
\end{aligned}
$$

Thus using the knowledge of quadratic function, we have:

$$
\begin{aligned}
||\Delta||_F^2 &= O_p(\rho_{p,T}K_T) + O_p(\log(p) + p\sqrt{\frac{\log(p)}{T}} + v_p^{-2(1+\varepsilon)}) \\
&+ O_p\left(\left(\left(\sqrt{\frac{\log(p)}{T}}\right)\sqrt{p+D_p} + \rho_{p,T}\sqrt{D_p}\right)^2\right. \\
&= O_p(\rho_{p,T}K_T + \log(p) + p\sqrt{\frac{\log(p)}{T}} + v_p^{-2(1+\varepsilon)} + \frac{D_p\log(p)}{T} + \rho_{p,T}^2 D_p).
\end{aligned}
$$

Then by Assumption 3, we have $D_p\log(p)/T = o(p\sqrt{\log(p)/T})$ as $D_p \ll p\sqrt{T/\log(p)}$. Denote

$$
\Delta_{\Sigma_e} = \sqrt{(\rho_{p,T}K_T + \log(p) + p\sqrt{\frac{\log(p)}{T}} + v_p^{-2(1+\varepsilon)} + \rho_{p,T}^2 D_p)/p}.
$$

We have $||\Delta||_F = O_p(p^{1/2}\Delta_{\Sigma_e})$. Notice that $||\widehat{\Sigma}_e - \Sigma_{e0}||_F$ is bounded by $||\Delta||_F$ as $\sigma_1(\Sigma_{e0})$ and $\sigma_1(\widehat{\Sigma}_e)$ are bounded. Under Assumption 3 and 5, we have:

$$
||\widehat{\Sigma}_e - \Sigma_{e0}||_F/\sqrt{p} = O_p(\Delta_{\Sigma_e}) = o_p(1). \tag{S4.6}
$$

Step II: This Step is similar to Bai and Liao (2016). We denote $R = (\widehat{B} - B_0)^\top \widehat{\Sigma}_e^{-1} \widehat{B} (\widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B})^{-1}$. To bound the term R, we need to bound $(I_r - R)^\top (I_r - R) - I_r$ and $B_0^\top \Sigma_{e0}^{-1} B_0 - (I_r - R) \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B} (I_r - R)^\top$.

First, we bound $B_0^\top \Sigma_{e0}^{-1} B_0 - (I_r - R) \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B} (I_r - R)^\top$, we have:

$$
\begin{aligned}
& B_0^\top \Sigma_{e0}^{-1} B_0 - (I_r - R) \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B} (I_r - R)^\top \\
&= B_0^\top \Sigma_{e0}^{-1} B_0 - B_0^\top \widehat{\Sigma}_e^{-1} \widehat{B} (\widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B})^{-1} \widehat{B}^\top \widehat{\Sigma}_e^{-1} B_0 \\
&= B_0^\top (\Sigma_{e0}^{-1} - \widehat{\Sigma}_e^{-1}) B_0 + B_0^\top \widehat{\Sigma}_e^{-1} B_0 - B_0^\top \widehat{\Sigma}_e^{-1} \widehat{B} (\widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B})^{-1} \widehat{B}^\top \widehat{\Sigma}_e^{-1} B_0 \\
&= O_p(p Q_2(\widehat{B}, \widehat{\Sigma}_e)) + B_0^\top (\Sigma_{e0}^{-1} - \widehat{\Sigma}_e^{-1}) B_0.
\end{aligned}
$$

The third equality can be obtained by noticing that the matrix $B_0^\top \widehat{\Sigma}_e^{-1} B_0 - B_0^\top \widehat{\Sigma}_e^{-1} \widehat{B} (\widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B})^{-1} \widehat{B}^\top \widehat{\Sigma}_e^{-1} B_0$ is semi-positive definite and hence its spectral norm is bounded by its trace. The term $B_0^\top (\Sigma_{e0}^{-1} - \widehat{\Sigma}_e^{-1}) B_0 = o_p(p)$ by Lemma 4. Thus we only need to bound $p Q_2(\widehat{B}, \widehat{\Sigma}_e)$. Remind (S4.3), we have

$$
\begin{aligned}
& p Q_2(\widehat{B}, \widehat{\Sigma}_e) \\
& \leq O_p(\log(p) + p \sqrt{\frac{\log(p)}{T}}) - (p Q_1(\widehat{\Sigma}_e) - p Q_1(\Sigma_{e0})) + p Q_2(J \widehat{B} + \mathbf{1}_p \gamma^\top, \Sigma_{e0}). \\
& \leq O_p(\log(p) + p \sqrt{\frac{\log(p)}{T}} + \rho_{p,T} K_T + \frac{(p + D_p) \log(p)}{T} + \rho_{p,T}^2 D_p + \upsilon_p^{-2(1+\varepsilon)}) \\
& = o_p(p).
\end{aligned}
$$

The second inequality above can be directly obtained from Lemma 3 that

$$
\begin{aligned}
&(pQ_1(\widehat{\Sigma}_e) - pQ_1(\Sigma_{e0}))\\
&\geq c_4\|\Delta\|_F^2 - \left(O_p\left(\sqrt{\frac{\log(p)}{T}}\right)\sqrt{p+D_p} + \rho_{p,T}\sqrt{D_p}\right)\|\Delta\|_F - 2\rho_{p,T}K_T.\\
&\geq O_p(\frac{(p+D_p)\log(p)}{T} + \rho_{p,T}^2 D_p) - 2\rho_{p,T}K_T.
\end{aligned}
$$

Thus we have:

$$
B_0^\top \Sigma_{e0}^{-1} B_0 - (I_r - R)\widehat{B}^\top \widehat{\Sigma}_e^{-1}\widehat{B}(I_r - R)^\top = o_p(p).
$$

Now we bound the $(I_r - R)^\top (I_r - R) - I_r$. To bound this term, we need the first order condition. For given $(J_p\widehat{B}, \mathbf{1}_p)$, we denote

$$
\widehat{B} = (J_p\widehat{B}, \mathbf{1}_p)\widehat{W} := (J_p\widehat{B}, \mathbf{1}_p)\begin{pmatrix} I_r \\ \mathbf{1}_p^\top \widehat{B}/p \end{pmatrix}.
$$

Then we have:

$$
\begin{aligned}
&(\widehat{W}, \widehat{\Sigma}_e, \widehat{\Omega}, \widehat{\alpha})\\
&\qquad = \operatorname{argmin}_{\{(W,\Sigma_e,\Omega,\alpha):((J_p\widehat{B},\mathbf{1}_p)W,\Sigma_e,\Omega,\alpha)\in\Xi_\delta\}}[L_Y((J_p\widehat{B}, \mathbf{1}_p)W, \Sigma_e)\\
&\qquad\quad + L_A((J_p\widehat{B}, \mathbf{1}_p)W, \Omega, \alpha) + P_T(\Sigma_e)].
\end{aligned}
$$

Denote

$$
\begin{aligned}
\Xi_{\delta,1} = \{&(W, \Sigma_e, \Omega, \alpha) : \delta_1^{-1/2} \leq \sigma_r((J_p\widehat{B}, \mathbf{1}_p)W/\sqrt{p}) \leq \sigma_r((J_p\widehat{B}, \mathbf{1}_p)W/\sqrt{p}) \leq \delta_1^{1/2},\\
&\delta_2^{-1/2} \leq \sigma_r((J_p\widehat{B}, \mathbf{0}_p)W/\sqrt{p}) \leq \sigma_r((J_p\widehat{B}, \mathbf{0}_p)W/\sqrt{p}) \leq \delta_2^{1/2},\\
&\max_{i,j}|(J_p\widehat{B}\Omega\widehat{B}^T J_p + \mathbf{1}_p\alpha^\top + \alpha\mathbf{1}_p^\top)_{ij}| \leq \delta_3,\\
&\text{and } \max\{||\Sigma_e||_1, ||\Sigma_e^{-1}||_1, ||\Sigma_e||_2, ||\Sigma_e^{-1}||_2\} \leq \delta_4\}.
\end{aligned}
$$

It is easy to verify that

$$(\widehat{W}, \widehat{\Sigma}_e, (\widehat{W}\widehat{\Omega}\widehat{W}^\top)_{[1:r,1:r]}, \widehat{\alpha}) = \text{argmin}_{\{(W,\Sigma_e,\Omega_1,\alpha)\in\Xi_{\delta,1}\}} L_1(W, \Sigma_e, \Omega_1, \alpha)$$

$$:= \text{argmin}_{\{(W,\Sigma_e,\Omega_1,\alpha)\in\Xi_{\delta,1}\}} [L_Y((J_p\widehat{B}, \mathbf{1}_p)W, \Sigma_e)$$

$$+ L_A(J_p\widehat{B}, \Omega_1, \alpha) + P_T(\Sigma_e)],$$

where $X_{[1:r,1:r]} \in \mathbb{R}^{r\times r}$ and $(X_{[1:r,1:r]})_{ij} = X_{ij}$, for matrix $X \in \mathbb{R}^{(r+k_1)\times(r+k_2)}$, $k_1, k_2 \in \mathbb{N}$. Thus, we have:

$$\partial L_1/\partial W\Big|_{\widehat{W},\widehat{\Sigma}_e,(\widehat{W}\widehat{\Omega}\widehat{W}^\top)_{[1:r,1:r]},\widehat{\alpha}} = \partial L_Y((J_p\widehat{B}, \mathbf{1}_p)W, \Sigma_e)/\partial W\Big|_{W=\widehat{W},\Sigma_e=\widehat{\Sigma}_e} = \mathbf{0}_{(r+1)\times r}.$$

By noticing that $(J_p\widehat{B}, \mathbf{1}_p)\widehat{W} = \widehat{B}$, we have:

$$\partial L_1/\partial W\Big|_{\widehat{W},\widehat{\Sigma}_e,(\widehat{W}\widehat{\Omega}\widehat{W}^\top)_{[1:r,1:r]},\widehat{\alpha}}$$

$$= (J_p\widehat{B}, \mathbf{1}_p)^\top(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e)^{-1}(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e)^{-1}\widehat{B} = \mathbf{0}_{(r+1)\times r}. \tag{S4.7}$$

Then we have

$$\widehat{B}^\top(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e)^{-1}(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e)^{-1}\widehat{B} = \mathbf{0}_{r\times r}.$$

Using the Sherman–Morrison–Woodbury formula we have

$$\widehat{B}^\top(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e)^{-1}$$

$$= \widehat{B}^\top(\widehat{\Sigma}_e^{-1} - \widehat{\Sigma}_e^{-1}\widehat{B}(I_r + \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1})$$

$$= (I_r - \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B}(I_r + \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1})\widehat{B}^\top\widehat{\Sigma}_e^{-1}$$

$$= ((I_r + \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})(I_r + B^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1} - \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B}(I_r + \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1})\widehat{B}^\top\widehat{\Sigma}_e^{-1}$$

$$= (I_r + \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}.$$

Thus we have

$$(I_r + \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B})^{-1} \widehat{B}^\top \widehat{\Sigma}_e^{-1} (\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y) \widehat{\Sigma}_e^{-1} \widehat{B} (I_r + \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B})^{-1} = \mathbf{0}_{r\times r}.$$

Then we multiply both sides of the above equation by $I_r + \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B}$ and H to obtain

$$H \widehat{B}^\top \widehat{\Sigma}_e^{-1} (\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y) \widehat{\Sigma}_e^{-1} \widehat{B} H = \mathbf{0}_{r\times r},$$

where $H = (\widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B})^{-1}$. We first point out that $||H||_2 = O_p(p^{-1})$ and $||H||_F \leq \sqrt{r}||H||_2 = O_p(p^{-1})$, which are easy to verify.

Under identification condition $T^{-1}FF^\top = I_r$ and $T^{-1}\sum_{t=1}^T f_t = 0$. We expand the S_y to obtain

$$\begin{aligned}
&- H \widehat{B}^\top \widehat{\Sigma}_e^{-1} (\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y) \widehat{\Sigma}_e^{-1} \widehat{B} H \\
&= -H \widehat{B}^\top \widehat{\Sigma}_e^{-1} (\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_e + \bar{e}\bar{e}^\top - B_0 B_0^\top - B_0 F \mathcal{E}^\top / T - \mathcal{E} F^\top B_0^\top / T) \widehat{\Sigma}_e^{-1} \widehat{B} H \\
&= H \widehat{B}^\top \widehat{\Sigma}_e^{-1} (S_e - \bar{e}\bar{e}^\top - \widehat{\Sigma}_e) \widehat{\Sigma}_e^{-1} \widehat{B} H + (I_r - R)^\top (I_r - R) \\
&+ (I_r - R)^\top F \mathcal{E}^\top \widehat{\Sigma}_e^{-1} \widehat{B} H / T + H \widehat{B}^\top \widehat{\Sigma}_e^{-1} \mathcal{E} F (I_r - R)/T - I_r = \mathbf{0}_{r\times r},
\end{aligned}$$

where $\bar{e} = \sum_{t=1}^T e_t / T$. Similar to the argument in Bai and Liao (2016) in Lemma A.5, we have $(I_r - R)^\top (I_r - R) - I_r = o_p(1)$. Combine $(I_r - R)^\top (I_r - R) - I_r = o_p(1)$ with $B_0^\top \Sigma_{e0}^{-1} B_0 - (I_r - R) \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B} (I_r - R)^\top = o_p(p)$ where the eigenvalues of $B_0^\top \Sigma_{e0}^{-1} B_0$ are distinct. We directly use Lemma 8 to obtain that $R = o_p(1)$, which is similar to (Bai and Liao, 2016).

Step III: Finally we use R to bound the $||\widehat{B}-B_0||_F$:

$$
\begin{aligned}
&Q_2(\widehat{B},\widehat{\Sigma}_e)+tr(\frac{1}{p}R\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B}R^\top)\\
&=p^{-1}tr(B_0^\top\widehat{\Sigma}_e^{-1}B_0-B_0^\top\widehat{\Sigma}_e^{-1}\widehat{B}(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}B_0)\\
&+p^{-1}tr((\widehat{B}-B_0)^\top\widehat{\Sigma}_e^{-1}\widehat{B}(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}(\widehat{B}-B_0))\\
&=p^{-1}[tr(B_0^\top\widehat{\Sigma}_e^{-1}B_0-B_0^\top\widehat{\Sigma}_e^{-1}\widehat{B}(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}B_0)\\
&+tr(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B}-B_0^\top\widehat{\Sigma}_e^{-1}\widehat{B}-\widehat{B}^\top\widehat{\Sigma}_e^{-1}B_0)+tr(B_0^\top\widehat{\Sigma}_e^{-1}\widehat{B}(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}B_0)]\\
&=p^{-1}tr(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B}-B_0^\top\widehat{\Sigma}_e^{-1}\widehat{B}-\widehat{B}^\top\widehat{\Sigma}_e^{-1}B_0+B_0^\top\widehat{\Sigma}_e^{-1}B_0)\\
&=tr\{\frac{1}{p}(\widehat{B}-B_0)^\top\widehat{\Sigma}_e^{-1}(\widehat{B}-B_0)\}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\frac{1}{p}\sigma_{\min}(\widehat{\Sigma}_e^{-1})||\widehat{B}-B_0||_F^2 &\le tr\{\frac{1}{p}(\widehat{B}-B_0)^\top\widehat{\Sigma}_e^{-1}(\widehat{B}-B_0)\}\\
&=Q_2(\widehat{B},\widehat{\Sigma}_e)+tr(\frac{1}{p}R\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B}R^\top)\\
&\le Q_2(\widehat{B},\widehat{\Sigma}_e)+\frac{1}{p}||R||_F^2||\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B}||_2=o_p(1).
\end{aligned}
$$

As $\sigma_{\min}(\widehat{\Sigma}_e^{-1})$ is bounded away from 0, we have $p^{-1}||\widehat{B}-B_0||_F^2=o_p(1)$. Hence $\min_{OO^\top=O^\top O=I_r}p^{-1}||\widehat{B}O-B_0||_F^2\le p^{-1}||\widehat{B}-B_0||_F^2=o_p(1)$.

Then we consider $\frac{1}{pT}||\widehat{B}\widehat{F} - B_0F||_F^2$. Notice that

$$\begin{aligned}\frac{1}{pT}||\widehat{B}\widehat{F} - B_0F||_F^2 &= \frac{1}{pT}||\widehat{B}\widehat{F} - \widehat{B}F + \widehat{B}F - B_0F||_F^2 \\ &\leq \frac{2}{pT}||\widehat{B}(\widehat{F} - F)||_F^2 + \frac{2}{pT}||(\widehat{B} - B_0)F||_F^2 \\ &\leq \frac{2}{pT}||\widehat{B}||_F^2||\widehat{F} - F||_F^2 + \frac{2}{pT}||F||_F^2||\widehat{B} - B_0||_F^2.\end{aligned}$$

We have $||\widehat{B}||_F^2 \leq r||\widehat{B}||_2^2 \leq r\delta_1 p = O_p(p)$ and $\frac{1}{T}||F||_F^2 = \frac{1}{T}tr(FF^\top) = tr(I_r) = O_p(1)$.

We have already proved that $||\widehat{B} - B_0||_F^2 = o_p(p)$. Now we consider $||\widehat{F} - F||_F^2$. using the result in Bai and Liao (2016) that $\widehat{f}_t - f_t = -R^\top f_t + (\widehat{B}^\top \widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}(e_t - \bar{e})$, we have:

$$\begin{aligned}\frac{1}{T}||\widehat{F} - F||_F^2 &= \frac{1}{T}\sum_{t=1}^{T}|| - R^\top f_t + (\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}(e_t - \bar{e})||_F^2 \\ &\leq \frac{2}{T}\sum_{t=1}^{T}(||R^\top f_t||_2^2 + ||(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}(e_t - \bar{e})||_F^2) = \mathcal{A}_1 + \mathcal{A}_2.\end{aligned}$$

The term $\mathcal{A}_1 \leq 2||R||_2^2\frac{1}{T}\sum_{t=1}^{T}||f_t||_2^2 = o_p(1)$. Now we consider the term

$\mathcal{A}_2$, as J_T is a projection matrix:

$$\begin{aligned}\mathcal{A}_2 &= 2T^{-1}||(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}\mathcal{E}J_T||_F^2 \\ &\leq 2T^{-1}||(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1}\mathcal{E}||_F^2 \\ &\leq 2T^{-1}||(\widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}||_F^2||\widehat{B}^\top\widehat{\Sigma}_e^{-1}\mathcal{E}||_F^2 \\ &\leq O_p((p^2T)^{-1})(||(\widehat{B}^\top\widehat{\Sigma}_e^{-1} - B_0^\top\Sigma_{e0}^{-1})\mathcal{E}||_F^2 + ||B_0^\top\Sigma_{e0}^{-1}\mathcal{E}||_F^2) \\ &= O_p((p^2T)^{-1})(\mathcal{A}_{21} + \mathcal{A}_{22}).\end{aligned}$$

Notice that

$$\begin{aligned}E\mathcal{A}_{22} &= Etr(B_0^\top\Sigma_{e0}^{-1}\mathcal{E}\mathcal{E}^\top\Sigma_{e0}^{-1}B_0) \\ &= tr(\Sigma_{e0}^{-1}B_0B_0^\top\Sigma_{e0}^{-1}E\mathcal{E}\mathcal{E}^\top) \\ &\leq ||E\mathcal{E}\mathcal{E}^\top||_2\sqrt{\text{rank}(\Sigma_{e0}^{-1}B_0B_0^\top\Sigma_{e0}^{-1})}||\Sigma_{e0}^{-1}B_0B_0^\top\Sigma_{e0}^{-1}||_F \\ &\leq rT||\Sigma_{e0}||_2||\Sigma_{e0}^{-1}B_0B_0^\top\Sigma_{e0}^{-1}||_2 = O(pT).\end{aligned}$$

We have $\mathcal{A}_{22} = O_p(pT)$. Then consider that

$$\mathcal{A}_{21} \leq 2||(\widehat{B}^\top - B_0^\top)\widehat{\Sigma}_e^{-1}\mathcal{E}||_F^2 + 2||B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})\mathcal{E}||_F^2 = \mathcal{A}_{211} + \mathcal{A}_{212}.$$

Notice that $E||\mathcal{E}||_F^2 = \sum_{i\leq p,t\leq T} E(e_{it}^2) = O(pT)$ and $||\widehat{B} - B_0||_F^2 = o_p(p)$, we have $\mathcal{A}_{211} \leq O_p(||\widehat{B} - B_0||_F^2||\mathcal{E}||_F^2) = o_p(p^2T)$. Now we turn to $\mathcal{A}_{212}$. Notice that

$$\mathcal{A}_{212} = O_p(||B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})\mathcal{E}||_2^2),$$

as $B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})\mathcal{E}$ is a low rank matrix. Thus we only need to bound

$||B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})\mathcal{E}||_2$. Notice that

$$B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})\mathcal{E} = B_0^\top\widehat{\Sigma}_e^{-1}(\Sigma_{e0} - \widehat{\Sigma}_e)\Sigma_{e0}^{-1}\mathcal{E}.$$

Remind u_i is the ith column of $B_0^\top\widehat{\Sigma}_e^{-1}$ and denote w_i is the ith row vector of $\Sigma_{e0}^{-1}\mathcal{E}$. We directly use Lemma 1 to obtain that $\max_{i\leq p}|T^{-1}\sum_{t=1}^T e_{it}^2| = O_p(1+\sqrt{\log(p)/T}) = O_p(1)$. Notice that $||\Sigma_{e0}^{-1}||_1$ and $||\widehat{\Sigma}_e^{-1}||_1$ are uniformly bounded, we have $\max_{i\leq p}||w_i||_2^2 = O_p(T)$. Then we have:

$$\begin{aligned}
&||B_0^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})\mathcal{E}||_2 \\
&\leq \sum_{i,j}|\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}|||u_i||_2||w_j||_2 \\
&\leq \max_{i\leq p}||u_i||_2 \max_{i\leq p}||w_i||_2(\sum_{J_L}|\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}| + \sum_{J_U}|\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}|) \\
&= o_p(p\sqrt{T}),
\end{aligned}$$

where we can directly obtain that $\sum_{J_L}|\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}| + \sum_{J_U}|\widehat{\Sigma}_{e,ij} - \Sigma_{e0,ij}| = o_p(p)$ in the proof of Lemma 4. Thus we have $a_2 = o_p(p^2T)$. Thus

$$\mathcal{A}_2 = O_p((p^2T)^{-1})(\mathcal{A}_{21} + \mathcal{A}_{22}) = O_p((p^2T)^{-1})(o_p(p^2T) + O_p(pT)) = o_p(1).$$

Hence we have $\min_{OO^\top=O^\top O=I_r} T^{-1}||O\widehat{F} - F||_F^2 \leq T^{-1}||\widehat{F} - F||_F^2 = \mathcal{A}_1 + \mathcal{A}_2 = o_p(1)$ and $(pT)^{-1}||\widehat{B}\widehat{F} - B_0F||_F^2 = o_p(1)$.

Part II: Now consider $\lambda \ll d_{p,T}$, we have:

$$L(\widehat{B}, \widehat{\Sigma}_e, \widehat{\Omega}, \widehat{\alpha}) \leq L(B_0, \Sigma_{e0}, \Omega_0, \alpha_0).$$

Thus

$$Q_1(\widehat{\Sigma}_e) + Q_2(\widehat{B}, \widehat{\Sigma}_e) + Q_3(\widehat{B}, \widehat{\Sigma}_e) + \frac{\lambda}{pT} L_A(\widehat{B}, \widehat{\Omega}, \widehat{\alpha})$$
$$\leq Q_1(\Sigma_{e0}) + Q_2(B_0, \Sigma_{e0}) + Q_3(B_0, \Sigma_{e0}) + \frac{\lambda}{pT} L_A(B_0, \Omega_0, \alpha_0).$$

Using Lemma 6, we have:

$$\sup_{(B,\Sigma_e,\Omega,\alpha)\in\Xi_\delta} (\frac{\lambda}{pT} L_A(B, \Omega, \alpha) - \frac{\lambda}{pT} L_A(B_0, \Omega_0, \alpha_0))^- \ll \max(\frac{\log(p)}{p}, \sqrt{\frac{\log(p)}{T}}).$$

Thus

$$Q_1(\widehat{\Sigma}_e) - Q_1(\Sigma_{e0}) + Q_2(\widehat{B}, \widehat{\Sigma}_e)$$
$$\leq -(\frac{\lambda}{pT} L_A(\widehat{B}, \widehat{\Omega}, \widehat{\alpha}) - \frac{\lambda}{pT} L_A(B_0, \Omega_0, \alpha_0)) + O_p(\frac{\log(p)}{p} + \sqrt{\frac{\log(p)}{T}})$$
$$\leq \sup_{(B,\Sigma_e,\Omega,\alpha)\in\Xi_\delta} (\frac{\lambda}{pT} L_A(B, \Omega, \alpha) - \frac{\lambda}{pT} L_A(B_0, \Omega_0, \alpha_0))^- + O_p(\frac{\log(p)}{p} + \sqrt{\frac{\log(p)}{T}}).$$

$$Q_1(\widehat{\Sigma}_e) - Q_1(\Sigma_{e0}) + Q_2(\widehat{B}, \widehat{\Sigma}_e) \leq O_p(\frac{\log(p)}{p} + \sqrt{\frac{\log(p)}{T}}).$$

This means that the effect of L_A can be absorbed into $O_p(\frac{\log(p)}{p} + \sqrt{\frac{\log(p)}{T}})$. Meanwhile, notice that $H\widehat{B}^\top \widehat{\Sigma}_e^{-1}(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1}\widehat{B}H = 0$ is still true. Thus we can totally imitate the proof in Bai and Liao (2016) to obtain the same convergence result. We don't repeat this procedure. We should point out that when $\sigma_r(\Omega_0) = 0$, this convergence result still holds for $\lambda \ll d_{p,T}$, as the result in Lemma 6 holds.

By the Lemma 7, notice that $\min\{v_p^{2\varepsilon}, \log^{-\varepsilon/(1+\varepsilon)}(p)\}d_{p,T} \ll d_{p,T}$. We directly obtain the consistency results for these estimators for any nonrandom positive $\lambda < +\infty$ which may vary as p, T changing.

S.5 Proof of Theorem 3 and Theorem S.1

We denote $B_1 = (J_p\widehat{B}, \mathbf{1}_p)$ and $B_2 = \widehat{\Sigma}_e^{-1/2} B_1$. Consider $\min_{OO^\top = O^\top O = I_r} ||\widehat{B}O - B_0||_F$, which has the convergence rate same to $\min_{OO^\top = O^\top O = I_r} ||\widehat{\Sigma}_e^{-1/2}\widehat{B}O - \widehat{\Sigma}_e^{-1/2}B_0||_F$. The latter can be controlled by the following sum of two parts:

$$||\widehat{\Sigma}_e^{-1/2}\widehat{B}O - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0||_F + ||P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0 - \widehat{\Sigma}_e^{-1/2}B_0||_F.$$

We divide the proof into three Steps. In Step I, we use Lemma 5 to bound the second term. In Step II, we use the first order condition to bound the first term. In Step III, we use $\min_{OO^\top = O^\top O = I_r} ||\widehat{B}O - B_0||$ to bound the $||\Sigma_Y - \Sigma_{Y0}||_{\Sigma_{Y0}}$. Then we detail the proof:

Step I: We bound the term $||P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0 - \widehat{\Sigma}_e^{-1/2}B_0||_F$. Using Lemma 5, we have:

$$\begin{aligned}||P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0 - \widehat{\Sigma}_e^{-1/2}B_0||_F &= \min_X ||\widehat{\Sigma}_e^{-1/2}B_0 - \widehat{\Sigma}_e^{-1/2}B_1X||_F \\ &\leq \sigma_{\max}(\widehat{\Sigma}_e^{-1/2}) \min_X ||B_0 - B_1X||_F \\ &\leq \delta_4^{1/2}||(I_p - P_{B_1})B_0||_F \\ &= \delta_4^{1/2}||(I_p - P_{B_1})J_pB_0||_F \\ &= \delta_4^{1/2}||(I_p - P_{J_p\widehat{B}} - P_{(I_p-P_{J_p\widehat{B}})\mathbf{1}_p})J_pB_0||_F \\ &= \delta_4^{1/2}||(I_p - P_{J_p\widehat{B}})J_pB_0||_F = O_p(v_p^{-1}).\end{aligned} \tag{S5.8}$$

Now we explain some equality in (S5.8). Notice that $P_{B_1}\mathbf{1}_p = \mathbf{1}_p$, we have $(I_p - P_{B_1})B_0 = (I_p - P_{B_1})J_pB_0 + (I_p - P_{B_1})(\mathbf{1}_p\mathbf{1}_p^\top/p)B_0 = (I_p - P_{B_1})J_pB_0$. Thus $\delta_4^{1/2}||(I_p-P_{B_1})B_0||_F = \delta_4^{1/2}||(I_p-P_{B_1})J_pB_0||_F$. Meanwhile notice that $(I_p - P_{J_p\widehat{B}})\mathbf{1}_p = \mathbf{1}_p$ and $P_{\mathbf{1}_p}J_p\widehat{B} = \mathbf{0}_{p\times r}$, which can be easily conducted by $J_p\mathbf{1}_p = \mathbf{0}_p$, we have $\delta_4^{1/2}||(I_p - P_{J_p\widehat{B}} - P_{(I_p-P_{J_p\widehat{B}})\mathbf{1}_p})J_pB_0||_F = \delta_4^{1/2}||(I_p - P_{J_p\widehat{B}})J_pB_0||_F$.

Step II: We consider the term $||\widehat{\Sigma}_e^{-1/2}\widehat{B}O - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0||_F$. To bound this term, we need to bound $||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0B_0^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F$ such that we can directly use Lemma 9. We have:

$$||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0B_0^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F - ||P_{B_2}||_F$$

$$- 2||P_{B_2}\widehat{\Sigma}_e^{-1/2}T^{-1}B_0F\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F - ||P_{B_2}\widehat{\Sigma}_e^{-1/2}T^{-1}\mathcal{E}J_T\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F$$

$$= ||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0B_0^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F - \mathcal{R}_1 - \mathcal{R}_2 - \mathcal{R}_3$$

$$\leq ||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} + P_{B_2}\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F.$$

First, we bound $||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} + P_{B_2}\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F$. To bound this term, we need the first order condition:

$$(J_p\widehat{B}, \mathbf{1}_p)^\top(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e)^{-1}(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e)^{-1}\widehat{B} = \mathbf{0}_{(r+1)\times r}.$$

Using the Sherman–Morrison–Woodbury formula we have $(J_p\widehat{B}, \mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1} + \widehat{\Sigma}_e^{-1}\widehat{B}(I_r + \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1}\widehat{B}^\top\widehat{\Sigma}_e^{-1})(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1}\widehat{B}(I_r + \widehat{B}^\top\widehat{\Sigma}_e^{-1}\widehat{B})^{-1} = \mathbf{0}_{(r+1)\times r}$. Notice that

$$\widehat{B}^\top\widehat{\Sigma}_e^{-1}(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1}\widehat{B} = \mathbf{0}_{r\times r},$$

after a simple calculation, we have:

$$(J_p\widehat{B}, \mathbf{1}_p)^\top\widehat{\Sigma}_e^{-1}(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1}\widehat{B} = \mathbf{0}_{(r+1)\times r}.$$

That is

$$B_2^\top\widehat{\Sigma}_e^{-1/2}(\widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1}\widehat{B} = \mathbf{0}_{(r+1)\times r}.$$

Remind that $\widehat{\Sigma}_e^{-1/2}\widehat{B} = \widehat{\Sigma}_e^{-1/2}B_1\widehat{W} = B_2\widehat{W}$, we have

$$B_2^\top P_{B_2}(B_2\widehat{W}\widehat{W}^\top B_2^\top + \widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2})P_{B_2}B_2\widehat{W} = \mathbf{0}_{(r+1)\times r}.$$

Denote $U = (B_2^\top B_2)^{1/2}\widehat{W}$ and $C = (B_2^\top B_2)^{-1/2} B_2^\top \widehat{\Sigma}_e^{-1/2}(S_y - \widehat{\Sigma}_e)\widehat{\Sigma}_e^{-1/2} B_2 (B_2^\top B_2)^{-1/2}$, we obtain that

$$UU^\top U - CU = \mathbf{0}_{(r+1)\times r}. \tag{S5.9}$$

To build the connection between $\widehat{B}\widehat{B}^\top$ and $UU^\top$. We first verify that $(B_2^\top B_2)^{1/2} X (B_2^\top B_2)^{1/2}$ share the same $r+1$ eigenvalues with $B_2 X B_2^\top$ for any symmetric matrix $X \in \mathbb{R}^{(r+1)\times(r+1)}$. Assume $B_2 = Q_{p\times p}\Lambda_{p\times(r+1)}V_{(r+1)\times(r+1)}$ is SVD decomposition. Then

$$(B_2^\top B_2)^{1/2} X (B_2^\top B_2)^{1/2} = V^\top (\Lambda^\top \Lambda)^{1/2} V X V^\top (\Lambda^\top\Lambda)^{1/2} V,$$

and

$$B_2 X B_2^\top = Q^\top \Lambda V X V^\top \Lambda Q = Q^\top \begin{pmatrix} (\Lambda^\top\Lambda)^{1/2} V X V^\top (\Lambda^\top\Lambda)^{1/2} & \mathbf{0}_{(r+1)\times(p-r-1)} \\ \mathbf{0}_{(r+1)\times(p-r-1)}^\top & \mathbf{0}_{(p-r-1)\times(p-r-1)} \end{pmatrix} Q.$$

Notice that V and Q are orthogonal. It is easy to verify they share same $r+1$ eigenvalues and the rest eigenvalues of $B_2^\top X B_2$ are all 0.

Assume $U = Q^*_{1,(r+1)\times r}\Lambda^*_{1,r\times r}V^*_{1,r\times r}$ is SVD decomposition. Notice that $\Lambda^*_{1,ii} \neq 0$ as $\text{rank}(U) = r$. Using result in (S5.9), We have $Q_1^*\Lambda_1^{*2} - CQ_1^* = \mathbf{0}_{(r+1)\times r}$, which implies $\Lambda^{*2}_{1,ii}$ is the k_ith eigenvalue of C and $Q^*_{1,\cdot i}$ is the eigenvector. Thus $UU^\top = Q^*\Lambda^{*2}Q^{*\top} = \sum_{i=1}^r \sigma_{k_i}(C)u_{k_i}(C)u_{k_i}^\top(C)$ where $u_i(C)$ is the i-th eigenvector and $k_1, \ldots, k_r$ are distinct value from $\{1, 2, \ldots, r+1\}$. Now we show that $\max_i\{k_i\} = r$. By Lemma 1, we have:

$$||\mathcal{E}\mathcal{E}^\top/T||_F \leq ||\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0}||_F + ||\Sigma_{e0}||_F = O_p(\sqrt{p} + p\sqrt{\log(p)/T}).$$

We should notice that $\sigma_r(UU^\top) = \sigma_r(\widehat{B}\widehat{B}^\top) \geq \delta_1^{-1}p$. Now we consider $\sigma_{r+1}(C)$. We have $\sigma_{r+1}(C) = \sigma_{r+1}(P_{B_2}\widehat{\Sigma}_e^{-1/2}(S_y - \widehat{\Sigma}_e)\widehat{\Sigma}_e^{-1/2}P_{B_2}) \leq 1 + \sigma_{r+1}(P_{B_2}\widehat{\Sigma}_e^{-1/2}S_y\widehat{\Sigma}_e^{-1/2}P_{B_2})$, which can be directly obtained by the fact $C = (B_2^\top B_2)^{1/2}(B_2^\top B_2)^{-1}B_2^\top\widehat{\Sigma}_e^{-1/2}(S_y - \widehat{\Sigma}_e)\widehat{\Sigma}_e^{-1/2}B_2(B_2^\top B_2)^{-1}(B_2^\top B_2)^{1/2}$ and $P_{B_2}\widehat{\Sigma}_e^{-1/2}(S_y - \widehat{\Sigma}_e)\widehat{\Sigma}_e^{-1/2}P_{B_2} = B_2(B_2^\top B_2)^{-1}B_2^\top\widehat{\Sigma}_e^{-1/2}(S_y - \widehat{\Sigma}_e)\widehat{\Sigma}_e^{-1/2}B_2(B_2^\top B_2)^{-1}B_2^\top$. Now we consider $\sigma_{r+1}(P_{B_2}\widehat{\Sigma}_e^{-1/2}S_y\widehat{\Sigma}_e^{-1/2}P_{B_2})$. Notice that

$$S_y = (B_0F + \mathcal{E})J_T(B_0F + \mathcal{E})^\top/T.$$

Thus

$$\begin{aligned}
\sigma_{r+1}(P_{B_2}S_yP_{B_2}) &= \sigma_{r+1}^2(P_{B_2}\widehat{\Sigma}_e^{-1/2}(B_0F + \mathcal{E})J_T/\sqrt{T}) \\
&\leq \sigma_{r+1}^2(P_{B_2}\widehat{\Sigma}_e^{-1/2}(B_0F + \mathcal{E})/\sqrt{T})\sigma_1^2(J_T) \\
&= \sigma_{r+1}^2(P_{B_2}\widehat{\Sigma}_e^{-1/2}(B_0F + \mathcal{E})/\sqrt{T}) \\
&\leq \sigma_1^2(P_{B_2}\widehat{\Sigma}_e^{-1/2}\mathcal{E}/\sqrt{T}) \\
&\leq ||P_{B_2}\widehat{\Sigma}_e^{-1/2}\mathcal{E}\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F/T \\
&= O_p(\sqrt{p} + p\sqrt{\log(p)/T}).
\end{aligned}$$

The forth inequality can be directly obtained by Weyl's inequality. Thus for $\log(p) \ll T$, we have $\sigma_r(UU^\top) \gg \sigma_{r+1}(C)$. And thus we have $UU^\top - C = \sigma_{r+1}(C)u_{r+1}(C)u_{r+1}^\top(C)$.

We have:

$$
\begin{aligned}
&||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} + P_{B_2}\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F \\
&= \sqrt{\sum_{i=1}^{r+1}\sigma_i^2(\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} + P_{B_2}\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}P_{B_2})} \\
&= \sqrt{\sum_{i=1}^{r+1}\sigma_i^2(UU^\top - C)} = ||UU^\top - C||_F \leq ||P_{B_2}\widehat{\Sigma}_e^{-1/2}\mathcal{E}\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F/T + 1.
\end{aligned}
$$

The second equality is obtained by the fact that $\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2}+P_{B_2}\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}P_{B_2} = B_2(\widehat{W}\widehat{W}^\top - (B_2^\top B_2)^{-1}B_2^\top\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}B_2(B_2^\top B_2)^{-1})B_2^\top$ share the same first $r+1$ eigenvalues with $UU^\top - C = (B_2^\top B_2)^{1/2}(\widehat{W}\widehat{W}^\top - (B_2^\top B_2)^{-1}B_2^\top\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}B_2(B_2^\top B_2)^{-1})(B_2^\top B_2)^{1/2}$.

Then we have:

$$
\begin{aligned}
\Delta_B &:= ||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0B_0^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F \\
&= O_p(\mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + ||\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2} + P_{B_2}\widehat{\Sigma}_e^{-1/2}(\widehat{\Sigma}_e - S_y)\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F) \\
&= O_p(||P_{B_2}\widehat{\Sigma}_e^{-1/2}T^{-1}\mathcal{E}\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F + ||P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0F\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F + 1) \\
&= O_p(\mathcal{H}_1 + \mathcal{H}_2 + 1).
\end{aligned}
\tag{S5.10}
$$

To obtain the desired result, we need to give a more accurate rate of $\mathcal{H}_1$ and $\mathcal{H}_2$. We first consider P_{B_2}. It is easy to verify that $\sigma_{\max}(B_2) \asymp p^{1/2}$ and $\sigma_{\min}(B_2) \asymp p^{1/2}$. Thus

$$
P_{B_2}\widehat{\Sigma}_e^{-1/2} = B_2(B_2^\top B_2)^{-1}B_2^\top\widehat{\Sigma}_e^{-1/2} = O_p(1/\sqrt{p})\begin{pmatrix}(X_1^\top)^{-1} & 0 \\ 0 & 1\end{pmatrix}(J_p\widehat{B}X_1, \mathbf{1}_p)^\top\widehat{\Sigma}_e^{-1},
$$

where $X_1 = O_2\Lambda_2^{-1/2}O_3\Lambda_1^{1/2}O_1^\top$ and notations Λ_1, Λ_2, O_1, O_2 and O_3 from the proof of Lemma 5. Thus $\sigma_r(X_1) \geq \sigma_r(\Lambda_1^{1/2})\sigma_r(\Lambda_2^{-1/2}) \geq (m/\delta_2)^{1/2}$, which implies $X_1^{-1} = O_p(1)$. Then we have

$$P_{B_2}\widehat{\Sigma}_e^{-1/2} = O_p(1/\sqrt{p})(J_p\widehat{B}X_1 - J_pB_0, \mathbf{0}_p)\widehat{\Sigma}_e^{-1} + O_p(1/\sqrt{p})(J_pB_0, \mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1}) + O_p(1/\sqrt{p})(J_pB_0, \mathbf{1}_p)^\top\Sigma_{e0}^{-1} = \mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3. \tag{S5.11}$$

Thus, we can easily write

$$\begin{aligned}\mathcal{H}_1 &= ||(\mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3)\mathcal{E}\mathcal{E}^\top(\mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3)^\top/T||_F \\ &\leq ||\widehat{\Sigma}_e^{-1/2}||_2||\mathcal{M}_1 + \mathcal{M}_2||_2||\mathcal{E}\mathcal{E}^\top/T||_F + ||\mathcal{M}_3\mathcal{E}\mathcal{E}^\top(\mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3)^\top/T||_F \\ &\leq (||\widehat{\Sigma}_e^{-1/2}||_2 + ||\mathcal{M}_3||_2)(||\mathcal{M}_1||_2 + ||\mathcal{M}_2||_2)||\mathcal{E}\mathcal{E}^\top/T||_F + ||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_F^2,\end{aligned}$$

and

$$\begin{aligned}\mathcal{H}_2 &\leq ||\widehat{\Sigma}_e^{-1/2}||_2||B_0F\mathcal{E}^\top(\mathcal{M}_1 + \mathcal{M}_2 + \mathcal{M}_3)^\top/T||_F \\ &\leq ||\widehat{\Sigma}_e^{-1/2}||_2(||\mathcal{M}_1||_2 + ||\mathcal{M}_2||_2)||B_0F\mathcal{E}^\top/T||_F + ||\widehat{\Sigma}_e^{-1/2}||_2||B_0F\mathcal{E}^\top\mathcal{M}_3^\top/T||_F \\ &\leq ||\widehat{\Sigma}_e^{-1/2}||_2((||\mathcal{M}_1||_2 + ||\mathcal{M}_2||_2)||B_0F\mathcal{E}^\top/T||_F + ||B_0||_F||F||_F||\mathcal{M}_3\mathcal{E}||_F/T).\end{aligned}$$

To bound $\mathcal{H}_1$ and $\mathcal{H}_2$, we need to bound $||\mathcal{M}_1||_2$, $||\mathcal{M}_2||_2$, $||\mathcal{M}_3||_2$ and $||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_F$.

Using Lemma 5, when $\lambda \gg d_{p,T}$, we have

$$\sigma_1(\mathcal{M}_1) \leq O_p(1/\sqrt{p})\sigma_1(\Lambda_1^{1/2})||J_pB_0O_1\Lambda_1^{-1/2} - J_p\widehat{B}O_2\Lambda_2^{-1/2}O_3||_F = O_p(v_p^{-1}/\sqrt{p}).$$

As we assume the 2-norms of row vectors of B_0 are uniformly bounded by a constant, it is easy to verify the norms of $(J_pB_0, \mathbf{1}_p)$'s row vectors are also uniformly bounded by a constant. Thus

$$\sigma_1(\mathcal{M}_2) \leq O_p(p^{-1/2})\sqrt{||(J_pB_0, \mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})^2(J_pB_0, \mathbf{1}_p)||_2}.$$

We have:

$$(J_pB_0, \mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})^2(J_pB_0, \mathbf{1}_p) = (J_pB_0, \mathbf{1}_p)^\top\Sigma_{e0}^{-1}(\Sigma_{e0} - \widehat{\Sigma}_e)\widehat{\Sigma}_e^{-1}(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})(J_pB_0, \mathbf{1}_p).$$

Denote $(J_pB_0, \mathbf{1}_p)^\top\Sigma_{e0}^{-1} = (h_1, ..., h_p)$ and $(J_pB_0, \mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})\widehat{\Sigma}_e^{-1} = (l_1, ..., l_p)$. Notice that $||\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1}||_1 \leq ||\widehat{\Sigma}_e^{-1}||_1 + ||\Sigma_{e0}^{-1}||_1$ is bounded by a constant. We have $\sup_{i\leq p}||h_i||_2 = O(1)$ and $\sup_{i\leq p}||l_i||_2 = O_p(1)$. By Lemma 4, we have $\sum_{i,j}|\Delta_{1,ij}| = O_p(p\theta_{p,T})$. Thus

$$\begin{aligned}
||(J_pB_0, \mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})^2(J_pB_0, \mathbf{1}_p)||_2 &\leq \sum_{i,j}|\Delta_{1,ij}|||l_i||_2||h_j||_2 \\
&\leq \sup_{i\leq p}||l_i||_2 \sup_{j\leq p}||h_j||_2 \sum_{i,j}|\Delta_{1,ij}| \\
&= O_p(p\theta_{p,T}).
\end{aligned}$$

Thus we have $||\mathcal{M}_2||_2 = O_p(\sqrt{\theta_{p,T}})$.

Finally, we consider the $||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_F$. We have:

$$||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_F \leq r^{1/2}||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_2 = O_p(p^{-1/2}||(J_pB_0, \mathbf{1}_p)^\top\Sigma_{e0}^{-1}\mathcal{E}/\sqrt{T}||_F).$$

Now, we bound $||(J_pB_0,\mathbf{1}_p)^\top\Sigma_{e0}^{-1}\mathcal{E}/\sqrt{T}||_F$. We should notice that

$$
\begin{aligned}
&E||(J_pB_0,\mathbf{1}_p)^\top\Sigma_{e0}^{-1}\mathcal{E}/\sqrt{T}||_F^2\\
&=E(tr((J_pB_0,\mathbf{1}_p)^\top\Sigma_{e0}^{-1}T^{-1}\mathcal{E}\mathcal{E}^\top\Sigma_{e0}^{-1}(J_pB_0,\mathbf{1}_p)))\\
&=tr[(J_pB_0,\mathbf{1}_p)^\top\Sigma_{e0}^{-1}T^{-1}E(\mathcal{E}\mathcal{E}^\top)\Sigma_{e0}^{-1}(J_pB_0,\mathbf{1}_p)]\\
&=tr(\Sigma_{e0}^{-1}(J_pB_0,\mathbf{1}_p)(J_pB_0,\mathbf{1}_p)^\top\Sigma_{e0}^{-1}\Sigma_{e0})\\
&\le||\Sigma_{e0}||_2\sqrt{r+1}||\Sigma_{e0}^{-1}(J_pB_0,\mathbf{1}_p)(J_pB_0,\mathbf{1}_p)^\top\Sigma_{e0}^{-1}||_F\\
&\le||\Sigma_{e0}||_2\sqrt{r+1}||\Sigma_{e0}^{-1}||_2^2||(J_pB_0,\mathbf{1}_p)||_F^2=O(p).
\end{aligned}
$$

Thus

$$
||(J_pB_0,\mathbf{1}_p)^\top\Sigma_{e0}^{-1}\mathcal{E}/\sqrt{T}||_F=O_p(\sqrt{p}),
$$

and hence

$$
||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_F=O_p(1).
$$

Now we back to $\mathcal{H}_1$ and $\mathcal{H}_2$. It is easy to verify that $||\mathcal{M}_3||_2=O_p(1)$. Then we have:

$$
\begin{aligned}
\mathcal{H}_1&\le(||\widehat{\Sigma}_e^{-1/2}||_2+||\mathcal{M}_3||_2)(||\mathcal{M}_1||_2+||\mathcal{M}_2||_2)||T^{-1}\mathcal{E}\mathcal{E}^\top||_F+||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_F^2\\
&\le O_p(v_p^{-1}/\sqrt{p}+\sqrt{\theta_{p,T}})||T^{-1}\mathcal{E}\mathcal{E}^\top||_F+O_p(1)\\
&=O_p((v_p^{-1}/\sqrt{p}+\sqrt{\theta_{p,T}})(p\sqrt{\log(p)/T}+\sqrt{p}))+O_p(1),
\end{aligned}
$$

and

$$\begin{aligned}\mathcal{H}_2 &\leq ||\widehat{\Sigma}_e^{-1/2}||_2(||\mathcal{M}_1||_2 + ||\mathcal{M}_2||_2)||T^{-1}B_0F\mathcal{E}^\top||_F \\ &+ ||\widehat{\Sigma}_e^{-1/2}||_2||B_0||_F||T^{-1/2}F||_F||\mathcal{M}_3\mathcal{E}/\sqrt{T}||_F \\ &\leq O_p(v_p^{-1}/\sqrt{p} + \sqrt{\theta_{p,T}})||T^{-1}B_0F\mathcal{E}^\top||_F + O_p(\sqrt{p}) \\ &= O_p((v_p^{-1}/\sqrt{p} + \sqrt{\theta_{p,T}})p\sqrt{\log(p)/T}) + O_p(\sqrt{p}).\end{aligned}$$

It is easy to verify $v_p^{-1}/\sqrt{p} \ll \sqrt{\theta_{p,T}} \ll 1$. Thus

$$\Delta_B = O_p(\mathcal{H}_1 + \mathcal{H}_2 + 1) = O_p(p\sqrt{\theta_{p,T}\log(p)/T} + \sqrt{p}).$$

Now we bound the term $\min_{OO^\top=O^\top O=I_r} ||\widehat{\Sigma}_e^{-1/2}\widehat{B}O - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0||$. We directly use Lemma 9 to obtain that

$$\begin{aligned}&p^{-1}\min_{OO^\top=O^\top O=I_r} ||\widehat{\Sigma}_e^{-1/2}\widehat{B}O - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0||_F^2 \\ &= O_p(\Delta_B^2/(p\lambda_r(\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2}))) = O_p(\theta_{p,T}\log(p)/T + 1/p).\end{aligned}$$

Combined with (S5.8), we have

$$p^{-1}\min_{OO^\top=O^\top O=I_r} ||\widehat{B}O - B_0||_F^2 = O_p(\theta_{p,T}\log(p)/T + v_p^{-2}/p).$$

Step III: Now we consider the term $||\widehat{\Sigma}_Y - \Sigma_{Y0}||_{\Sigma_{Y0}}$.

$$||\widehat{\Sigma}_Y - \Sigma_{Y0}||_{\Sigma_{Y_0}} \leq ||\widehat{B}\widehat{B}^\top - B_0B_0^\top||_{\Sigma_{Y_0}} + ||\widehat{\Sigma}_e - \Sigma_{e0}||_{\Sigma_{Y_0}}.$$

Notice that

$$||\widehat{\Sigma}_e - \Sigma_{e0}||_{\Sigma_{Y_0}} = p^{-1/2}||\Sigma_{Y0}^{-1/2}(\widehat{\Sigma}_e - \Sigma_{e0})\Sigma_{Y0}^{-1/2}||_F \leq p^{-1/2}\sigma_{\max}(\Sigma_{Y0}^{-1})||\widehat{\Sigma}_e - \Sigma_{e0}||_F,$$

and $\sigma_{\max}(\Sigma_{Y0}^{-1}) \leq \sigma_{\min}^{-1}(B_0 B_0^\top + \Sigma_e) \leq c^{-1}$. As proved in Theorem 2, when $\lambda \gg d_{pT}$, $||\widehat{\Sigma}_e - \Sigma_{e0}||_{\Sigma_{Y_0}}$ is bounded by $O_p(\Delta_{\Sigma_e})$. Now we consider the term $||\widehat{B}\widehat{B}^\top - B_0 B_0^\top||_{\Sigma_{Y_0}}$.

$$||\widehat{B}\widehat{B}^\top - B_0 B_0^\top||_{\Sigma_{Y_0}} \leq ||(\widehat{B}O - B_0)(\widehat{B}O - B_0)^\top||_{\Sigma_{Y_0}} + 2||(\widehat{B}O - B_0)B_0^\top||_{\Sigma_{Y_0}},$$

for any orthogonal matrix O. We choose O which minimize $||\widehat{B}O - B_0||_F$.

The term

$$\begin{aligned} ||(\widehat{B}O - B_0)(\widehat{B}O - B_0)^\top||_{\Sigma_{Y_0}} &\leq p^{-1/2}\sigma_1(\Sigma_{Y_0}^{-1})||(\widehat{B}O - B_0)(\widehat{B}O - B_0)^\top||_F \\ &\leq O_p(p^{-1/2}||(\widehat{B}O - B_0)||_F^2) \\ &= O_p(\sqrt{p}\theta_{p,T}\log(p)/T + v_p^{-2}/\sqrt{p}). \end{aligned}$$

Now we consider $||(\widehat{B}O - B_0)B_0^\top||_{\Sigma_Y}$. We have

$$p^{-1/2}||\Sigma_{Y0}^{-1/2}(\widehat{B}O - B_0)B_0^\top \Sigma_{Y0}^{-1/2}||_F^2 = p^{-1/2}tr(\Sigma_{Y0}^{-1}(\widehat{B}O - B_0)B_0^\top \Sigma_{Y0}^{-1} B_0(\widehat{B}O - B_0)^\top).$$

Using the Sherman-Morrison-Woodbury formula we have

$$\Sigma_{Y0}^{-1} = \Sigma_{e0}^{-1} - \Sigma_{e0}^{-1} B_0 \left(I_r + B_0^\top \Sigma_{e0}^{-1} B_0\right)^{-1} B_0^\top \Sigma_{e0}^{-1}.$$

And thus

$$\begin{aligned} B_0^\top \Sigma_{Y0}^{-1} B_0 &= B_0^\top \Sigma_{e0}^{-1} B_0 \left(I_r + B_0^\top \Sigma_{e0}^{-1} B_0\right)^{-1} \left(I_r + B_0^\top \Sigma_{e0}^{-1} B_0 - B_0^\top \Sigma_{e0}^{-1} B_0\right) \\ &= \left(I_r + B_0^\top \Sigma_{e0}^{-1} B_0 - I_r\right) \left(I_r + B_0^\top \Sigma_{e0}^{-1} B_0\right)^{-1} \\ &= I_r - \left(I_r + B_0^\top \Sigma_{e0}^{-1} B_0\right)^{-1}. \end{aligned}$$

Combined with $\left(I_r + B_0^\top \Sigma_{e0}^{-1} B_0\right)^{-1} = O_p(1)$. We have $||B_0^\top \Sigma_{Y0}^{-1} B_0||_F = O_p(1)$. Then, we have

$$
\begin{aligned}
& p^{-1}||\Sigma_{Y0}^{-1/2}(\widehat{B}O - B_0)B_0^\top \Sigma_{Y0}^{-1/2}||_F^2 \\
& = p^{-1} tr(\Sigma_{Y0}^{-1}(\widehat{B}O - B_0)B_0^\top \Sigma_{Y0}^{-1} B_0(\widehat{B}O - B_0)^\top) \\
& \leq \sqrt{r} p^{-1}||\Sigma_{Y0}^{-1}||_2||(\widehat{B}O - B_0)B_0^\top \Sigma_{Y0}^{-1} B_0(\widehat{B}O - B_0)^\top||_F \\
& \leq \sqrt{r} p^{-1}||\Sigma_{Y0}^{-1}||_2||\widehat{B}O - B_0||_F^2||B_0^\top \Sigma_{Y0}^{-1} B_0||_F \\
& = O_p(\theta_{p,T}\log(p)/T + v_p^{-2}/p).
\end{aligned}
$$

Thus we have

$$
||\widehat{\Sigma}_Y - \Sigma_{Y0}||_{\Sigma_{Y0}} = O_p(\sqrt{\theta_{p,T}\log(p)/T} + \sqrt{p}\theta_{p,T}\log(p)/T + v_p^{-2}/\sqrt{p} + \Delta_{\Sigma_e}).
$$

Finally we absorb $\sqrt{\theta_{p,T}\log(p)/T}$ into Δ_{Σ_e} for desired result. Theorem 3 is a special case of this result, which can be proved by directly substituting $\log(p) \ll T \ll p^{4/5}$, $\max\{K_T, v_p^{-1}\} = O(1)$, and $D_p \asymp p$ into the convergence rate of Theorem S.1.

S.6 Proof of Theorem 4

The proof of Theorem 4 is totally similar to the proof of Theorem 3. We will give the rate of the term Δ_B in (S5.10) when $\widehat{\Sigma}_e$ is diagonal, which is controlled by $\mathcal{H}_1 = ||P_{B_2}\widehat{\Sigma}_e^{-1/2}T^{-1}\mathcal{E}\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F$ and $\mathcal{H}_2 = ||P_{B_2}\widehat{\Sigma}_e^{-1/2}T^{-1}B_0F\mathcal{E}^\top\widehat{\Sigma}_e^{-1/2}P_{B_2}||_F$. We first consider the term P_{B_2}. It is easy to verify that $\sigma_{\max}(B_2) \asymp p^{1/2}$ and $\sigma_{\min}(B_2) \asymp p^{1/2}$. Thus

$$P_{B_2}\widehat{\Sigma}_e^{-1/2} = B_2(B_2^\top B_2)^{-1}B_2^\top \widehat{\Sigma}_e^{-1/2} = O_p(1/\sqrt{p}) \begin{pmatrix} (X_1^\top)^{-1} & 0 \\ 0 & 1 \end{pmatrix} (J_p\widehat{B}X_1, \mathbf{1}_p)^\top \widehat{\Sigma}_e^{-1},$$

where we remind that $X_1 = O_2\Lambda_2^{-1/2}O_3\Lambda_1^{1/2}O_1^\top$. Thus $\sigma_r(X_1) \geq \sigma_r(\Lambda_1^{1/2})\sigma_r(\Lambda_2^{-1/2}) \geq (m/\delta_2)^{1/2}$, which implies $X_1^{-1} = O_p(1)$. We assume Σ_{e0} is diagonal in Theorem 4. Then we have

$$P_{B_2}\widehat{\Sigma}_e^{-1/2} = O_p(1/\sqrt{p})(J_p\widehat{B}X_1 - J_pB_0, \mathbf{0}_p)\widehat{\Sigma}_e^{-1} + O_p(1/\sqrt{p})(J_pB_0, \mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1})$$
$$+ O_p(1/\sqrt{p})(J_pB_0, \mathbf{1}_p)^\top \Sigma_{e0}^{-1} = \mathcal{D}_1 + \mathcal{D}_2 + \mathcal{D}_3. \tag{S6.12}$$

Then, we have

$$\mathcal{H}_1 = ||(\mathcal{D}_1 + \mathcal{D}_2 + \mathcal{D}_3)\mathcal{E}\mathcal{E}^\top(\mathcal{D}_1 + \mathcal{D}_2 + \mathcal{D}_3)^\top/T||_F$$
$$\leq (||\widehat{\Sigma}_e^{-1/2}||_2 + ||\mathcal{D}_3||_2)(||\mathcal{D}_1||_2 + ||\mathcal{D}_2||_2)||\mathcal{E}\mathcal{E}^\top/T||_F + ||\mathcal{D}_3\mathcal{E}/\sqrt{T}||_F^2,$$

and

$$\mathcal{H}_2 \leq ||\widehat{\Sigma}_e^{-1/2}||_2||T^{-1}B_0F\mathcal{E}^\top(\mathcal{D}_1 + \mathcal{D}_2 + \mathcal{D}_3)^\top||_F$$
$$\leq ||\widehat{\Sigma}_e^{-1/2}||_2(||\mathcal{D}_1||_2 + ||\mathcal{D}_2||_2)||B_0F\mathcal{E}^\top/T||_F + ||\widehat{\Sigma}_e^{-1/2}||_2||B_0F\mathcal{E}^\top\mathcal{D}_3^\top/T||_F$$
$$\leq ||\widehat{\Sigma}_e^{-1/2}||_2((||\mathcal{D}_1||_2 + ||\mathcal{D}_2||_2)||B_0F\mathcal{E}^\top/T||_F + ||B_0||_F||F||_F||\mathcal{D}_3\mathcal{E}||_F/T).$$

To bound $\mathcal{H}_1$ and $\mathcal{H}_2$, we need to bound $||\mathcal{D}_1||_2$, $||\mathcal{D}_2||_2$, $||\mathcal{D}_3||_2$ and $||\mathcal{D}_3\mathcal{E}/\sqrt{T}||_F$.

Using Lemma 5, when $\lambda \gg d_{p,T}$, we have

$$\sigma_1(\mathcal{D}_1) \leq O_p(1/\sqrt{p})\sigma_1(\Lambda_1^{1/2})||J_pB_0O_1\Lambda_1^{-1/2} - J_p\widehat{B}O_2\Lambda_2^{-1/2}O_3||_F = O_p(v_p^{-1}/\sqrt{p}).$$

As we assume the 2-norms of row vectors of B_0 are uniformly bounded by a constant, it is easy to verify the norms of $(J_pB_0, \mathbf{1}_p)$'s row vectors are also uniformly bounded by a constant. Thus

$$\begin{aligned}\sigma_1(\mathcal{D}_2) &\leq O_p(p^{-1/2})\sqrt{tr((J_pB_0,\mathbf{1}_p)^\top(\widehat{\Sigma}_e^{-1}-(\Sigma_{e0})^{-1})^2(J_pB_0,\mathbf{1}_p))} \\ &\leq O_p(p^{-1/2})\sqrt{O(1)\sum_{i=1}^{p}((\widehat{\Sigma}_{e,ii})^{-1}-(\Sigma_{e0,ii})^{-1})^2}.\end{aligned} \tag{S6.13}$$

We need to bound $p^{-1}\sum_{i=1}^{p}((\widehat{\Sigma}_{e,ii})^{-1}-(\Sigma_{e0,ii})^{-1})^2$. Denote $p\widetilde{Q}_1(\Sigma_e) = \log|\Sigma_e| + tr(S_e\Sigma_e^{-1})$, we have:

$$L(\widehat{B},\widehat{\Sigma}_e,\widehat{\Omega},\widehat{\alpha}) \leq L(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\Sigma_{e0},\widehat{\Omega},\widehat{\alpha}),$$

and then

$$(p\widetilde{Q}_1(\widehat{\Sigma}_e)-p\widetilde{Q}_1(\Sigma_{e0}))+(pQ_2(\widehat{B},\widehat{\Sigma}_e)-pQ_2(J_p\widehat{B}+\mathbf{1}_p\gamma^\top,\Sigma_{e0})) \leq O_p(\log(p)+p\sqrt{\frac{\log(p)}{T}}),$$

where the γ from Proof of Theorem 2. Similar to argument in Proof of Theorem 2, for $\lambda \gg d_{p,T}$ we have:

$$p\widetilde{Q}_1(\widehat{\Sigma}_e)-p\widetilde{Q}_1(\Sigma_{e0})+pQ_2(\widehat{B},\widehat{\Sigma}_e) \leq O_p(\log(p)+p\sqrt{\frac{\log(p)}{T}}+v_p^{-2}). \tag{S6.14}$$

We imitate the proof in Bai and Liao (2016). Denote $f(t) = -\log|\Sigma_{e0}^{-1} +$

$t\Delta_2| + tr(S_e(\Sigma_{e0}^{-1} + t\Delta_2))$, where $\Delta_2 = \widehat{\Sigma}_e^{-1} - \Sigma_{e0}^{-1}$ is diagonal. We have:

$$
\begin{aligned}
|f'(0)| &= |tr(\Delta_2(S_e - (\Sigma_{e0}^{-1} + t\Delta_2)^{-1}))|_{t=0}| \\
&= |tr((S_e - \Sigma_{e0})\Delta_2)| \leq \max_{i \leq p} |S_{e,ii} - \Sigma_{e0,ii}| \sum_i |\Delta_{2,ii}| \\
&\leq O_p(\sqrt{\log(p)/T}) \sum_i |\Delta_{2,ii}| \leq O_p(\sqrt{p\log(p)/T})||\Delta_2||_F,
\end{aligned}
$$

and

$$
\begin{aligned}
f''(t) &= tr((\Sigma_{e0}^{-1} + t\Delta_2)^{-1}\Delta_2(\Sigma_{e0}^{-1} + t\Delta_2)^{-1}\Delta_2) \\
&= vec(\Delta_2)^\top[(\Sigma_{e0}^{-1} + t\Delta_2)^{-1} \otimes (\Sigma_{e0}^{-1} + t\Delta_2)^{-1}]vec(\Delta_2).
\end{aligned}
$$

As $\lambda_{\max}(\Sigma_{e0}^{-1} + t\Delta_2) = \lambda_{\max}((1-t)\Sigma_{e0}^{-1} + t\widehat{\Sigma}_e^{-1}) \leq \delta_4 + c_1^{-1}$ for $t \in [0,1]$, we have $\lambda_{\min}((\Sigma_{e0}^{-1} + t\Delta_2)^{-1}) \geq (\delta_4 + c_1^{-1})^{-1}$. Thus there exists a constant d such that

$$
\begin{aligned}
f''(t) &\geq \lambda_{\min}((\Sigma_{e0}^{-1} + t\Delta_2)^{-1} \otimes (\Sigma_{e0}^{-1} + t\Delta_2)^{-1})||vec(\Delta_2)||_2^2 \\
&= \lambda_{\min}^2((\Sigma_{e0}^{-1} + t\Delta_2)^{-1})||vec(\Delta_2)||_2^2 \geq d||\Delta_2||_F^2,
\end{aligned}
$$

for $t \in [0,1]$. Thus

$$
p\widetilde{Q}_1(\widehat{\Sigma}_e) - p\widetilde{Q}_1(\Sigma_{e0}) = f(1) - f(0) = f'(0) + f''(\xi) \geq -O_p(\sqrt{p\log(p)/T})||\Delta_2||_F + d||\Delta_2||_F^2, \tag{S6.15}
$$

where $\xi \in (0,1)$. From (S6.14), we have $p\widetilde{Q}_1(\widehat{\Sigma}_e) - p\widetilde{Q}_1(\Sigma_{e0}) \leq O_p(\log(p) + p\sqrt{\frac{\log(p)}{T}} + v_p^{-2})$ by $pQ_2(\widehat{B}, \widehat{\Sigma}_e) \geq 0$. Thus

$$
p^{-1}||\Delta_2||_F^2 = p^{-1}\sum_{i=1}^p ((\widehat{\Sigma}_{e,ii})^{-1} - \Sigma_{e0,ii}^{-1})^2 = \Delta_{\Sigma,dg}^2 := O_p(\log(p)/p + \sqrt{\log(p)/T} + v_p^{-2}/p). \tag{S6.16}
$$

By (S6.13), we have

$$\mathcal{D}_2 = O_p(\Delta_{\Sigma,dg}).$$

Finally, we can totally imitate the Proof of Theorem 3 to obtain that

$$||\mathcal{D}_3\mathcal{E}/\sqrt{T}||_F = O_p(1).$$

It is easy to verify that $||\mathcal{D}_3||_2 = O_p(1)$. Then we have:

$$\begin{aligned}\mathcal{H}_1 &\leq (||\widehat{\Sigma}_e^{-1/2}||_2 + ||\mathcal{D}_3||_2)(||\mathcal{D}_1||_2 + ||\mathcal{D}_2||_2)||T^{-1}\mathcal{E}\mathcal{E}^\top||_F + ||\mathcal{D}_3\mathcal{E}/\sqrt{T}||_F^2 \\ &\leq O_p(v_p^{-1}/\sqrt{p} + \Delta_{\Sigma,dg})||T^{-1}\mathcal{E}\mathcal{E}^\top||_F + O_p(1) \\ &= O_p((v_p^{-1}/\sqrt{p} + \Delta_{\Sigma,dg})(p\sqrt{\log(p)/T} + \sqrt{p})) + O_p(1),\end{aligned}$$

and

$$\begin{aligned}\mathcal{H}_2 &\leq ||\widehat{\Sigma}_e^{-1/2}||_2((||\mathcal{D}_1||_2 + ||\mathcal{D}_2||_2)||T^{-1}B_0F\mathcal{E}^\top||_F + ||B_0||_F||F||_F||\mathcal{D}_3\mathcal{E}||_F/T) \\ &\leq O_p(v_p^{-1}/\sqrt{p} + \Delta_{\Sigma,dg})||T^{-1}B_0F\mathcal{E}^\top||_F + O_p(\sqrt{p}) \\ &= O_p((v_p^{-1}/\sqrt{p} + \Delta_{\Sigma,dg})p\sqrt{\log(p)/T}) + O_p(\sqrt{p}).\end{aligned}$$

Now we simplify $(v_p^{-1}/\sqrt{p} + \Delta_{\Sigma,dg})(p\sqrt{\log(p)/T} + \sqrt{p})$, we have:

$$\begin{aligned}&(v_p^{-1}/\sqrt{p} + \Delta_{\Sigma,dg})(p\sqrt{\log(p)/T} + \sqrt{p}) \\ &= O(\left[(\log(p)/p)^{1/2} + (\log(p)/T)^{1/4} + v_p^{-1}/p^{1/2}\right]\left[p(\log(p)/T)^{1/2} + \sqrt{p}\right]) \\ &= O(p^{1/2}\log(p)/T^{1/2} + p(\log(p)/T)^{3/4} + p^{1/2}(\log(p)/T)^{1/2}v_p^{-1}) \\ &+ O(log(p)^{1/2} + p^{1/2}(\log(p)/T)^{1/4} + v_p^{-1}) \\ &= O(p^{1/2}\log(p)/T^{1/2} + p(\log(p)/T)^{3/4} + p^{1/2}v_p^{-1} + p^{1/2}) \\ &= O(p^{1/2}\log(p)/T^{1/2} + p(\log(p)/T)^{3/4} + p^{1/2}v_p^{-1}).\end{aligned}$$

Then, use the same method in proof of Theorem 3, we have:

$$\begin{aligned}
&\min_{OO^\top=O^\top O=I_r} p^{-1/2}||\widehat{\Sigma}_e^{-1/2}\widehat{B}O - P_{B_2}\widehat{\Sigma}_e^{-1/2}B_0||_F \\
&= p^{-1/2}O_p(\Delta_B/\sqrt{\lambda_r(\widehat{\Sigma}_e^{-1/2}\widehat{B}\widehat{B}^\top\widehat{\Sigma}_e^{-1/2})}) \\
&= p^{-1}O_p(\mathcal{H}_1+\mathcal{H}_2+1) \\
&= O_p(\log(p)/\sqrt{pT}+(\log(p)/T)^{3/4}+v_p^{-1}/\sqrt{p}).
\end{aligned}$$

Using a similar method, we obtain the result same as (S5.8) . Thus we have

$$\min_{OO^\top=O^\top O=I_r} p^{-1}||\widehat{B}O - B_0||_F^2 = O_p(\frac{\log^2(p)}{pT}+\left(\frac{\log(p)}{T}\right)^{3/2}+\frac{1}{pv_p^2}).$$

S.7 Proof of Proposition 1

The main idea to prove the desired result is using Theorem 3 in Bai and Li (2012). This result suggests that when $\lambda = 0$, we may have $\widehat{B}_{ML} - B_0 = \mathcal{E}F^\top/T + o_p(\sqrt{p/T})$. In this proof, we need to use theorems and lemmas in Bai and Li (2012). Thus we divide the proof into three Steps. Step I is introduced to prove assumptions in Bai and Li (2012) are satisfied. In Step II, we prove that given F, $\widehat{B}_{ML} - B_0 = \mathcal{E}F^\top/T + o_p(\sqrt{p/T})$ is true. In Step III, we prove the desired result base on the fact $\widehat{B}_{ML} - B_0 = \mathcal{E}F^\top/T + o_p(\sqrt{p/T})$. In Bai and Li (2012), factors f_t are assumed to be fixed. Thus we need to consider the convergence result given factors F. In the following content, without specified the O_p and o_p are all conditioning on $\{f_t\}_{t=1}^T$. We use $\Pr_{\cdot|\mathrm{F}}$ and $E_{\cdot|F}$ to denote the conditional probability and expectation of $\{f_t\}_{t=1}^T$.

Step I: In this step, we prove that given bounded $\{f_t\}_{t=1}^T$, Assumptions A-E in Bai and Li (2012) are satisfied under Assumptions 1-3, 4(1'), 6 and 7 in this article such that we can directly use lemmas and theorems in Bai and Li (2012). Assumptions A, B, C.1-3 and D and E are obviously satisfied. The bound of 8-th order moment in C.1 can be obtained by the uniform exponential tail. We denote $E_{\cdot|F}(e_{it}^8) = E(e_{it}^8) \leq C_e^8$ as $\{e_t\}_{t=1}^T$ and $\{f_t\}_{t=1}^T$ are independent. Thus for all $n \leq 8$, $(E|e_{it}|^n)^{1/n} \leq C_e$.

To show that C.4 is satisfied, notice that $E_{\cdot|F}[e_{it}e_{is}] = E[e_{it}e_{is}]$. Under Assumption 6, we have:

$$|E(e_{it}e_{is})| \leq \rho(|s-t|)\sqrt{E(|e_{it}|^2)E(|e_{is}|^2)} \leq C_e^2 \exp(-a_3'|s-t|^{r_3}).$$

Denote $\rho_{st} = C_e^2 \exp(-a_3'|s-t|^{r_3})$, we have:

$$T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\rho_{st} \leq 2\sum_{d=0}^{\infty}\rho_{0d} \leq \sum_{d=0}^{\infty} 2C_e^2 \exp(-a_3' d^{r_3'}).$$

This sum is convergence. Thus $T^{-1}\sum_{t=1}^T\sum_{s=1}^T \rho_{st}$ is bounded by a constant that only depends on C_e, a_3', r_3'.

To verify Lemma C.5 in (Bai and Li, 2012), we set $g(f_t, e_t) = e_{it}e_{jt} - E(e_{it}e_{jt})$ and apply Lemma 10. The moment condition in Lemma 10 can be easily verified by a uniform exponential tail. Thus we prove that A-E in

(Bai and Li, 2012) are all satisfied.

Step II: In this step, we will prove $\widehat{B}_{ML} - B_0 = \mathcal{E}F^\top/T + o_p(\sqrt{p/T})$. First, we transform the $\partial L/\partial B$ to obtain the formula of $\widehat{B}_{ML} - B_0$. We have:

$$\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{B}_{ML}\widehat{B}_{ML}^\top + \widehat{\Sigma}_{e,ML} - S_y) = \mathbf{0}_{r\times p}.$$

Thus we have

$$\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\widehat{B}_{ML}(\widehat{B}_{ML} - B_0)^\top + \widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{B}_{ML} - B_0)B_0^\top$$
$$+ \widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{\Sigma}_{e,ML} - \Sigma_{e0}) - \widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}B_0 F\mathcal{E}^\top/T - \widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\mathcal{E}F^\top B_0^\top/T$$
$$- \widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0}) + \widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\bar{e}\bar{e}^\top = \mathbf{0}_{r\times p}.$$

Remind $H = (\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\widehat{B}_{ML})^{-1}$, we have

$$(\widehat{B}_{ML} - B_0)^\top = -H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{B}_{ML} - B_0)B_0^\top$$
$$- H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{\Sigma}_{e,ML} - \Sigma_{e0}) + H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}B_0 F\mathcal{E}^\top/T + H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\mathcal{E}F^\top B_0^\top/T$$
$$+ H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0}) - H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\bar{e}\bar{e}^\top$$
$$= \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 + \mathcal{J}_4 + \mathcal{J}_5 + \mathcal{J}_6. \tag{S7.17}$$

Now we show that $\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_4, \mathcal{J}_5$ and $\mathcal{J}_6$ are all $o_p(\sqrt{p/T})$. We directly use lemmas in (Bai and Li, 2012). The proof of Theorem 3 in Bai and Li (2012) implies that $H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{B}_{ML} - B_0) = O_p((pT)^{-1/2} + p^{-1} + T^{-1})$. Thus

$$\mathcal{J}_1 = -H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{B}_{ML} - B_0)B_0^\top = O_p(T^{-1/2} + p^{-1/2} + p^{1/2}T^{-1}) = o_p(\sqrt{p/T}).$$

For $\mathcal{J}_2$, notice that $p \gg T$, it is easy to verify $\mathcal{J}_2 = O_p(||H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\widehat{\Sigma}_{e,ML} - \Sigma_{e0})||_2) \leq O_p(||H||_2||\widehat{B}_{ML}||_2||\widehat{\Sigma}_{e,ML}^{-1}||_2||\widehat{\Sigma}_{e,ML} - \Sigma_{e0}||_2) = O_p(p^{-1/2}) = o_p(\sqrt{p/T})$, as $||\widehat{\Sigma}_{e,ML} - \Sigma_{e0}||_2 \leq ||\widehat{\Sigma}_{e,ML}||_2 + ||\Sigma_{e0}||_2 = O_p(1)$.

Using Lemma C.1 in (Bai and Li, 2012), we have $H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\mathcal{E}F^\top/T = O_p((pT)^{-1/2} + T^{-1})$. Thus

$$\mathcal{J}_4 = H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}\mathcal{E}F^\top B_0^\top/T = O_p(T^{-1/2} + p^{1/2}T^{-1}) = o_p(\sqrt{p/T}).$$

Now we bound $\mathcal{J}_5$. We first consider $||H\widehat{B}_{ML}^\top \widehat{\Sigma}_{e,ML}^{-1}(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F$. It is bounded by

$$||HB_0^\top \Sigma_e^{*-1}(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F + ||H(\widehat{B}_{ML} - B_0)^\top \widehat{\Sigma}_{e,ML}^{-1}(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F$$
$$+ ||HB_0^\top(\widehat{\Sigma}_{e,ML}^{-1} - \Sigma_e^{*-1})(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F = \mathcal{K}_1 + \mathcal{K}_2 + \mathcal{K}_3,$$

where $\Sigma_e^* \in \mathbb{R}^{p\times p}$ and $\Sigma_{e,ij}^* = \Sigma_{e0,ij}1_{i=j}$. For $\mathcal{K}_1$, we have:

$$\begin{aligned}\mathcal{K}_1 &\leq ||H||_F||B_0^\top \Sigma_e^{*-1}(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F \\ &\leq O_p(p^{-1})\sqrt{\sum_{j=1}^p ||\sum_{i=1}^p\sum_{t=1}^T(\Sigma_{e0,ii})^{-1}b_{0i}[e_{it}e_{jt} - E(e_{it}e_{jt})]/T||_2^2}.\end{aligned}$$

Under Assumption 7 (3) we have

$$E(\sum_{j=1}^p ||\sum_{i=1}^p\sum_{t=1}^T(\Sigma_{e0,ii})^{-1}b_{0i}[e_{it}e_{jt} - E(e_{it}e_{jt})]||_2^2)$$
$$= E(pT\sum_{j=1}^p E||(pT)^{-1/2}\sum_{i=1}^p\sum_{t=1}^T(\Sigma_{e0,ii})^{-1}b_{0i}[e_{it}e_{jt} - E(e_{it}e_{jt})]||_2^2) \quad \leq Kp^2T.$$

And thus

$$\sqrt{\sum_{j=1}^p ||\sum_{i=1}^p\sum_{t=1}^T(\Sigma_{e0,ii})^{-1}b_{0i}[e_{it}e_{jt} - E(e_{it}e_{jt})]/T||_2^2} = O_p(pT^{-1/2}),$$

$$\mathcal{K}_1 = O_p(T^{-1/2}).$$

For term $\mathcal{K}_2$, we have $\mathcal{K}_2 \leq O_p(p^{-1})||\widehat{B}_{ML} - B_0||_F||(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F$.

The Assumption C.5 in Bai and Li (2012) implies that

$$\begin{aligned}
E||\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0}||_F^2 &= \sum_{i=1}^p\sum_{j=1}^p E[(\sum_{t=1}^T(e_{it}e_{jt} - E(e_{it}e_{jt}))/T)^2] \\
&\leq p^2T^{-1}\max_{i,j} E[(T^{-1/2}\sum_{t=1}^T(e_{it}e_{jt} - E(e_{it}e_{jt})))^2] \\
&\leq p^2T^{-1}\max_{i,j}(E[(T^{-1/2}\sum_{t=1}^T(e_{it}e_{jt} - E(e_{it}e_{jt})))^4])^{1/2} \\
&= O(p^2/T),
\end{aligned}$$

which implies that $||\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0}||_F = O_p(p/\sqrt{T})$. Notice that $p^{-1}||\widehat{B}_{ML} - B_0||_F^2 = O_p(T^{-1} + p^{-2})$ directly by Theorem 2 in (Bai and Li, 2012), we have:

$$\mathcal{K}_2 = O_p(\sqrt{p}/T + \sqrt{1/(pT)}).$$

For term $\mathcal{K}_3$, we have

$$\begin{aligned}
\mathcal{K}_3^2 &\leq ||H||_F^2||B_0^\top(\widehat{\Sigma}_{e,ML}^{-1} - \Sigma_e^{*-1})||_2^2||(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F^2 \\
&= O_p(p^{-2}\sum_{i=1}^p(\widehat{\Sigma}_{e,ML,ii}^{-1} - (\Sigma_{e0,ii})^{-1})^2||(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F^2) \\
&= O_p((p^{-3} + p^{-1}T^{-1})||(\mathcal{E}\mathcal{E}^\top/T - \Sigma_{e0})||_F^2).
\end{aligned}$$

The second row is directly from

$$||B_0^\top(\widehat{\Sigma}_{e,ML}^{-1}-\Sigma_e^{*-1})||_2^2 \leq tr(B_0^\top(\widehat{\Sigma}_{e,ML}^{-1}-\Sigma_e^{*-1})^2B_0) = O_p(1)\sum_{i=1}^p(\widehat{\Sigma}_{e,ML,ii}^{-1}-(\Sigma_{e0,ii})^{-1})^2,$$

as the row vectors of B_0 are uniformly bounded. Then the bound of $p^{-1}\sum_{i=1}^{p}(\widehat{\Sigma}_{e,ML,ii}^{-1}-(\Sigma_{e0,ii})^{-1})^2$ is given as

$$
\begin{aligned}
p^{-1}\sum_{i=1}^{p}(\widehat{\Sigma}_{e,ML,ii}^{-1}-(\Sigma_{e0,ii})^{-1})^2 &= p^{-1}||\widehat{\Sigma}_{e,ML}^{-1}-\Sigma_e^{*-1}||_F^2 \\
&= p^{-1}||\widehat{\Sigma}_{e,ML}^{-1}(\widehat{\Sigma}_{e,ML}-\Sigma_e^{*})\Sigma_e^{*-1}||_F^2 \\
&\le p^{-1}||\widehat{\Sigma}_{e,ML}^{-1}||_2^2||\widehat{\Sigma}_{e,ML}-\Sigma_e^{*}||_F^2||\Sigma_e^{*-1}||_2^2 \\
&= O_p(T^{-1}+p^{-2}).
\end{aligned}
$$

Remind $||\mathcal{E}\mathcal{E}^\top/T-\Sigma_{e0}||_F=O_p(p/\sqrt{T})$, we have $\mathcal{K}_3=O_p(\sqrt{p}/T+\sqrt{1/(pT)})$.

Thus we have

$$\mathcal{J}_5=O_p(\mathcal{K}_1+\mathcal{K}_2+\mathcal{K}_3)=o_p(\sqrt{p/T}).$$

Finally we consider

$$\mathcal{J}_6=H\widehat{B}_{ML}^\top\widehat{\Sigma}_{e,ML}^{-1}\bar{e}\bar{e}^\top=O_p(p^{-1/2}||\bar{e}||_2^2)=O_p(\sqrt{p}/T)=o_p(\sqrt{p/T}).$$

The third equality can be obtained by noticing that

$$
\begin{aligned}
E(||\bar{e}||_2^2) &= \sum_{i=1}^{p}E(\bar{e}_i)^2 \\
&= T^{-2}\sum_{i=1}^{p}\sum_{t=1}^{T}\sum_{s=1}^{T}E(e_{it}e_{is}) \\
&\le 2T^{-2}\sum_{i=1}^{p}\sum_{t=1}^{T}\sum_{s=1}^{\infty}C_e\rho(|s-t|)=O(p/T).
\end{aligned}
$$

Thus, we have

$$\widehat{B}_{ML}-B_0=(H\widehat{B}_{ML}^\top\widehat{\Sigma}_{e,ML}^{-1}B_0F\mathcal{E}^\top/T)^\top+o_p(\sqrt{p/T})=\mathcal{E}F^\top/T+o_p(\sqrt{p/T}),$$

as $H\widehat{B}_{ML}^\top\widehat{\Sigma}_{e,ML}^{-1}(B_0 - \widehat{B}_{ML}) = o_p(1)$ and $\mathcal{E}F^\top/T = O_p(\sqrt{p/T})$ by Lemma B.2 in Bai and Li (2012) and $||(H\widehat{B}_{ML}^\top\widehat{\Sigma}_{e,ML}^{-1}B_0)^{-1}||_2 = ||(I_r - o_p(1))^{-1}||_2 = O_p(1)$. We should notice that $\mathcal{J}_{1:6}$ and $(H\widehat{B}_{ML}^\top\widehat{\Sigma}_{e,ML}^{-1}(B_0 - \widehat{B}_{ML})F\mathcal{E}^\top/T)^\top$ are all low-rank matrices, which implies that the rate of their Frobenius norms is same to the rate of their 2-norms. Thus we have

$$\begin{aligned}
\min_{OO^\top=O^\top O=I_r} ||\widehat{B}_{ML}O - B_0||_F &= \min_{OO^\top=O^\top O=I_r} ||\widehat{B}_{ML} - B_0O||_F \\
&= \min_{OO^\top=O^\top O=I_r} ||(\widehat{B}_{ML} - B_0) - B_0(O - I)||_F \\
&\geq \min_{X\in\mathbb{R}^{r\times r}} ||(\widehat{B}_{ML} - B_0) - B_0X||_F \\
&= ||(I_p - P_{B_0})(\mathcal{E}F^\top/T + o_p(\sqrt{p/T}))||_F \\
&\geq ||\mathcal{E}F^\top/T||_F - ||P_{B_0}\mathcal{E}F^\top/T||_F - o_p(\sqrt{p/T}).
\end{aligned}$$

Notice that $||P_{B_0}\mathcal{E}F^\top/T||_F \leq ||P_{B_0}\mathcal{E}/\sqrt{T}||_F||F/\sqrt{T}||_F$ and combine with $E||P_{B_0}\mathcal{E}/\sqrt{T}||_F^2 = Etr(P_{B_0}T^{-1}\mathcal{E}\mathcal{E}^\top P_{B_0}) = tr(P_{B_0}\Sigma_{e0}) = O(1)$, we have $||P_{B_0}\mathcal{E}/\sqrt{T}||_F = O_p(1)$ and then we have $||P_{B_0}\mathcal{E}F^\top/T||_F = O_p(1) = o_p(\sqrt{p/T})$. Thus

$$\min_{OO^\top=O^\top O=I_r} ||\widehat{B}_{ML}O - B_0||_F \geq ||\mathcal{E}F^\top/T||_F - o_p(\sqrt{p/T}).$$

Step III: Denote $X_2 = p^{-1}||\mathcal{E}F^\top/\sqrt{T}||_F^2$. Recall that m_1 is defined in

Assumption 6, we consider $\zeta \in \mathbb{R}$

$$\mathrm{Pr}_{\cdot|\mathrm{F}}(\min_{\mathrm{OO}^\top=\mathrm{O}^\top\mathrm{O}=\mathrm{I_r}} \mathrm{T/p}||\widehat{\mathrm{B}}_{\mathrm{ML}}\mathrm{O}-\mathrm{B}_0||_\mathrm{F}^2 \geq \mathrm{m}_1\zeta)$$

$$\geq \mathrm{Pr}_{\cdot|\mathrm{F}}(\mathrm{X}_2 \geq 1.1\mathrm{m}_1\zeta) - \mathrm{Pr}_{\cdot|\mathrm{F}}(\min_{\mathrm{OO}^\top=\mathrm{O}^\top\mathrm{O}=\mathrm{I_r}} \mathrm{T/p}||\widehat{\mathrm{B}}_{\mathrm{ML}}\mathrm{O}-\mathrm{B}_0||_\mathrm{F}^2 - \mathrm{X}_2 \leq -0.1\mathrm{m}_1\zeta)$$

$$= \mathcal{C}_1 + \mathcal{C}_2.$$

Then, we will illustrate that $E_{\cdot|F}(X_2) \geq m_1$ almost surely and use the Paley–Zygmund inequality to give the lower bound of $\mathcal{C}_1$. Thus we need the first and second order moment of X_2. We have:

$$\begin{aligned}
E_{\cdot|F}(X_2) &= (pT)^{-1}E_{\cdot|F}(\sum_{i=1}^p ||(\sum_{t=1}^T f_t e_{it})||_F^2) \\
&= (pT)^{-1}E_{\cdot|F}(\sum_{i=1}^p\sum_{t=1}^T\sum_{s=1}^T tr(f_t f_s^\top e_{it}e_{is})) \\
&= (pT)^{-1}\sum_{i=1}^p\sum_{t=1}^T\sum_{s=1}^T tr(f_t f_s^\top cov(e_{it}, e_{is})) \\
&\geq m_1,
\end{aligned}$$

and

$$\begin{aligned}
E_{\cdot|F}(X_2^2) &= p^{-2}E_{\cdot|F}(\sum_{i=1}^p\sum_{j=1}^r(\sum_{t=1}^T f_{jt}e_{it}/\sqrt{T})^2)^2 \\
&\leq p^{-2}E_{\cdot|F}((\sum_{i=1}^p\sum_{j=1}^r 1^2)(\sum_{i=1}^p\sum_{j=1}^r(\sum_{t=1}^T f_{jt}e_{it}/\sqrt{T})^4)).
\end{aligned}$$

Denote $g(f_t, e_t) = f_{jt}e_{it}$. Notice that $E_{\cdot|F}(g(f_t, e_t)) = 0$ and notice that $\mathrm{Pr}_{\cdot|\mathrm{F}}(|\mathrm{f_{jt}e_{it}}| \geq \mathrm{s}) \leq \exp(-(\mathrm{s}/(\mathrm{a}_1\mathrm{K}))^{\mathrm{r}_1})$ almost surely as $||f_t||_2$ is uniformly bounded by K, we have $E_{\cdot|F}(z_t^6) \leq \int_0^\infty 6s^5\exp(-(s/(a_1K))^{r_1})ds$ is bounded by a constant doesn't depend on i, j, p, t, T and F. Directly use Lemma 10, we have $E_{\cdot|F}[((\sum_{t=1}^{T} f_{jt}e_{it})/\sqrt{T})^4]$ is uniformly bounded by a constant C_{e1} that doesn't depend on i, j, p, T and F. We have:

$$E_{\cdot|F}(X_2^2) \leq r^2 C_{e1}.$$

Thus for $\zeta < 1/1.1$, we have $\mathcal{C}_1 \geq \mathrm{Pr}_{\cdot|F}(X_2 \geq 1.1\zeta \mathrm{E}_{\cdot|F}(X_2)) \geq (1 - 1.1\zeta)^2 \mathrm{E}_{\cdot|F}^2(X_2)/\mathrm{E}_{\cdot|F}(X_2^2) \geq (1-1.1\zeta)^2 m_1^2/(r^2 C_{e1})$.

Now we bound $\mathcal{C}_2$. We denote $\mathcal{C} = \sqrt{T/p}\min_{OO^\top = O^\top O = I_r} ||\widehat{B}_{ML}O - B_0||_F$. By the fact $\mathcal{C} - \sqrt{X_2} \geq o_p(1)$ and $0 \leq \mathcal{C} = O_p(T^{1/2}(p^{-1} + T^{-1/2})) = O_p(1)$, we have $\mathcal{C}^2 - X_2 = (\mathcal{C} + \sqrt{X_2})(\mathcal{C} - \sqrt{X_2}) \geq o_p(1)$. Thus $\mathcal{C}_2$ converge to 0. Thus we have:

$$\begin{aligned}
&\liminf_{p,T\to\infty} \Pr(\min_{OO^\top = O^\top O = I_r} p^{-1}||\widehat{B}_{ML}O - B_0||_F^2 \geq T^{-1} m_1 \zeta) \\
&\geq E[\liminf_{p,T\to\infty} \Pr_{\cdot|F}(\min_{OO^\top = O^\top O = I_r} p^{-1}||\widehat{B}_{ML}O - B_0||_F^2 \geq T^{-1} m_1 \zeta)] \\
&\geq (1-1.1\zeta)^2 m_1^2/(r^2 C_{e1}).
\end{aligned}$$

For example, we can choose $\zeta = 0.5$, then $k_1 = 0.5m_1$ and $k_2 = 0.2025 m_1^2/(r^2 C_{e1})$ for the desired result.

S.8 Additional Algorithmic Details

We use cross-validation to select the tuning parameters λ and $\rho_{p,T}$, following Bai and Liao (2016). The index set $\{1, 2, \ldots, T\}$ is divided into K subsets of approximately equal size, denoted as $\{\mathcal{T}_k\}_{k \leq K}$, with minor adjustments if $T/K \notin \mathbb{Z}$. For each fold k, let $Y_{,-\mathcal{T}_k}$ denote the submatrix of Y obtained by excluding the columns indexed by $\mathcal{T}_k$, and $Y_{,\mathcal{T}_k}$ denote the submatrix of Y containing only the columns indexed by $\mathcal{T}_k$. Obtain the estimator, $\widehat{B}_{k,(\lambda,\rho_{p,T})}$ and $\widehat{\Sigma}_{e,k,(\lambda,\rho_{p,T})}$ using $Y_{,-\mathcal{T}_k}$, with fixed λ and $\rho_{p,T}$. The average loss function is given by

$$\mathcal{L}_{\lambda,\rho_{p,T}} = K^{-1}\sum_{k\leq K} L_{Y_{,\mathcal{T}_k}}(\widehat{B}_{k,(\lambda,\rho_{p,T})}, \widehat{\Sigma}_{e,k,(\lambda,\rho_{p,T})}),$$

where $L_{Y_{,\mathcal{T}_k}}(\widehat{B}_{k,(\lambda,\rho_{p,T})}$ is proportional to the negative log-likelihood of $Y_{,\mathcal{T}_k}$. The optimal values of λ and $\rho_{p,T}$ are chosen by minimizing $\mathcal{L}_{\lambda,\rho_{p,T}}$.

S.9 Details of Comparison Methods

In this subsection, we briefly introduce the PCA, POET, PML, TSM, PC-L, and ML-nL methods used for comparison.

(a) **PCA method (Bai, 2003):** The PCA method provides the estimators of the factor model by solving the following optimization problem:

$$(\widehat{B}^{PCA}, \widehat{F}^{PCA}) = \text{argmin}_{(B,F):F^\top F/T=I_r} ||Y - BF||_F^2.$$

Furthermore, the covariance matrix estimator of the idiosyncratic error component is given by $\widehat{\Sigma}_e^{PCA} = T^{-1}(Y - \widehat{B}^{PCA}\widehat{F}^{PCA})(Y - \widehat{B}^{PCA}\widehat{F}^{PCA})^\top$. Finally, the covariance matrix estimator of the series Y is given by $\widehat{\Sigma}_Y^{PCA} = \widehat{B}^{PCA}(\widehat{B}^{PCA})^\top + \widehat{\Sigma}_e^{PCA}$.

(b) **POET method (Fan et al., 2013):** The POET estimator of factor loadings matrix and factors are given by $\widehat{B}^{POET} = \widehat{B}^{PCA}, \widehat{F}^{POET} = \widehat{F}^{PCA}$. The estimator of Σ_{e0} is defined by $\widehat{\Sigma}_{e,ij}^{POET} = \mathcal{S}(\widehat{\Sigma}_{e,ij}^{PCA}, \tau_{ij} I_{i\neq j})$, where $\mathcal{S}(a,b) = \mathrm{sign}(a)(|a|-b)^+$, and τ_{ij} is the threshold. The threshold term $\tau_{ij} = C(1/\sqrt{p} + \sqrt{\log(p)/T})\widehat{\theta}_{ij}^{1/2}$, where $\widehat{\theta}_{ij} = T^{-1}\sum_{t=1}^{T}(\widehat{e}_{it}^{PCA}\widehat{e}_{jt}^{PCA} - \widehat{\Sigma}_{e,ij}^{PCA})^2$ and $\widehat{e}_{it}^{PCA} = Y_{it} - (\widehat{b}_i^{PCA})^\top \widehat{f}_t^{PCA}$. Then the covariance matrix estimator of the series Y is given by $\widehat{\Sigma}_Y^{POET} = \widehat{B}^{POET}(\widehat{B}^{POET})^\top + \widehat{\Sigma}_e^{POET}$.

(c) **PML method (Bai and Liao, 2016):** The PML estimator of the factor loadings B_0 and the covariance matrix Σ_{e0} are given by

$$(\widehat{B}^{PML}, \widehat{\Sigma}_e^{PML}) = \mathrm{argmin}_{(B,\Sigma_e)}\, L_Y(B,\Sigma_e) + P_T(\Sigma_e),$$

where $L_Y(B,\Sigma_e) = \log(\det(BB^\top + \Sigma_e)) + \mathrm{tr}(S_y(BB^\top + \Sigma_e)^{-1})$, $P_T(\Sigma_e) = \sum_{i\neq j} \rho_{p,T}|\Sigma_{e,ij}|$ and $\rho_{p,T}$ is the tuning parameter. The details of the algorithms for solving the above optimization problem can be found in Bai and Liao (2016). The covariance matrix estimator of the series Y is given by $\widehat{\Sigma}_Y^{PML} = \widehat{B}^{PML}(\widehat{B}^{PML})^\top + \widehat{\Sigma}_e^{PML}$, and the estimator of factor is given by $\widehat{F}^{PML} = ((\widehat{B}^{PML})^\top(\widehat{\Sigma}_e^{PML})^{-1}\widehat{B}^{PML})^{-1}(\widehat{B}^{PML})^\top(\widehat{\Sigma}_e^{PML})^{-1}Y$.

(d) **TSM method:** We analyze A using the project gradient descent algorithm to obtain the $\tilde{\Theta}_A$ as an estimator of $\Theta_{A0} = J_p B_0 \Omega_0 B_0^\top J_p + \alpha_0 \mathbf{1}_p^\top + \mathbf{1}_p \alpha_0^\top$ (Zhang et al., 2020).Then, we find $\widehat{\tilde{\Gamma}}$ such that $J_p \tilde{\Theta}_A J_p = \widehat{\tilde{\Gamma}} I_{\tilde{q}_1,\tilde{q}_2} \widehat{\tilde{\Gamma}}^\top$, where $\tilde{q}_1, \tilde{q}_2$ are the numbers of positive and negative eigenvalues of $\tilde{\Theta}_A$, respectively. The TSM estimator of B_0 and Σ_{e0} are given by

$$(\widehat{B}^{TSM}, \widehat{\Sigma}_e^{TSM}) = \operatorname{argmin}_{(B,\Sigma_e)} L_Y(B, \Sigma_e) + P_T(\Sigma_e), \text{ with } J_p B = \widehat{\bar{\Gamma}} \bar{W}, \text{ for some } \bar{W}.$$

The estimator of Σ_Y and F are $\widehat{\Sigma}_Y^{TSM} = \widehat{B}^{TSM}(\widehat{B}^{TSM})^\top + \widehat{\Sigma}_e^{TSM}$, and $\widehat{F}^{TSM} = ((\widehat{B}^{TSM})^\top (\widehat{\Sigma}_e^{TSM})^{-1} \widehat{B}^{TSM})^{-1} (\widehat{B}^{TSM})^\top (\widehat{\Sigma}_e^{TSM})^{-1} Y$, respectively.

(e) **PC-L method:** The PC-L method provides the estimators of the factor model by solving the following optimization problem:

$$(\widehat{B}^{PC-L}, \widehat{F}^{PC-L}) = \operatorname{argmin}_{(B,F):FF^\top/T=I_r} (pT)^{-1} ||Y - BF||_F^2 + p^{-1} \alpha^* tr(B^\top (D - A) B),$$

where D is a diagonal matrix with $D_{ii} = \sum_{j=1}^p A_{ij}$, and α^* is the tuning parameter. The covariance matrix estimator of the idiosyncratic error component is given by $\widehat{\Sigma}_e^{PC-L} = T^{-1}(Y - \widehat{B}^{PC-L}\widehat{F}^{PC-L})(Y - \widehat{B}^{PC-L}\widehat{F}^{PC-L})^\top$, and the covariance matrix estimator of the series Y is given by $\widehat{\Sigma}_Y^{PC-L} = \widehat{B}^{PC-L}(\widehat{B}^{PC-L})^\top + \widehat{\Sigma}_e^{PC-L}$.

(f) **ML-nL method:** The ML-nL method provides the estimators of the factor model based on the normalized Laplacian, Specifically,

$$(\widehat{B}^{ML-nL}, \widehat{\Sigma}_e^{ML-nL}) = \operatorname{argmin}_{(B,\Sigma_e)} L_Y(B, \Sigma_e) + \lambda_{nL} tr(B^\top (I_p - D_1^{-1/2} A D_1^{-1/2}) B) + P_T(\Sigma_e),$$

where λ_{nL} is the tuning parameter, D_1 is a diagonal matrix with $D_{1,ii} = \max\{D_{ii}, 1\}$. The estimator of the Σ_Y is given by $\widehat{\Sigma}_Y^{ML-nL} = \widehat{B}^{ML-nL}(\widehat{B}^{ML-nL})^\top + \widehat{\Sigma}_e^{ML-nL}$, and the estimator of factor is given by

$$\widehat{F}^{ML-nL} = ((\widehat{B}^{ML-nL})^\top (\widehat{\Sigma}_e^{ML-nL})^{-1} \widehat{B}^{ML-nL})^{-1} (\widehat{B}^{ML-nL})^\top (\widehat{\Sigma}_e^{ML-nL})^{-1} Y.$$

(g) **covar-based method:** We introduce how we use this method in real data analysis. We consider the following model:

$$Y_t = X_c\beta + B_0 f_t + e_t,$$

where $X_c \in \mathbb{R}^{p\times K}$ and $X_{c,ij} = 1$ if and only if $j = 1$ or i-th stock belong to $(j-1)$-th group for $j > 1$, K is the number of industries, B_0 is the factor loading matrix, f_t is the unobservable factor vector, and e_t is the idiosyncratic error vector with mean zero and covariance matrix Σ_e. The estimator $\widehat{\beta}$ is defined as $(X_c^\top X_c)X_c Y \mathbf{1}_T/T$. Then we apply PCA to $Y - X_c\widehat{\beta}$ to obtain $\widehat{B}^{cov-based}$ and $\widehat{F}^{cov-based}$. The estimator of Σ_{e0} is given by $\widehat{\Sigma}_{e,ij}^{cov-based} = T^{-1}(Y - X_c\widehat{\beta} - \widehat{B}^{cov-based}\widehat{F}^{cov-based})(Y - X_c\widehat{\beta} - \widehat{B}^{cov-based}\widehat{F}^{cov-based})^\top I_{i=j}$. Then the covariance matrix estimator of the series Y is given by $\widehat{\Sigma}_Y^{cov-based} = \widehat{B}^{cov-based}(\widehat{B}^{cov-based})^\top + \widehat{\Sigma}_e^{cov-based}$.

S.10 Additional Simulation Results

S.10.1 The simulation results of Example III

We provide the results of Example III in the simulation setting in Table S.1.

Table S.1: Simulation results of Example III. Each cell shows the mean×10 (standard error ×10).

T	Case	ME_B	ME_{Σ_Y}	ME_{Σ_e}	ME_F
300	PML	0.82(0.07)	2.04(0.08)	0.94(0.06)	2.09(0.09)
	Case (a)	0.61(0.05)	1.80(0.07)	0.94(0.06)	2.08(0.09)
	Case (b)	0.67(0.05)	1.87(0.07)	0.94(0.06)	2.09(0.09)
	Case (c)	0.75(0.06)	1.97(0.07)	0.94(0.06)	2.08(0.09)
500	PML	0.51(0.04)	1.56(0.06)	0.76(0.05)	2.07(0.07)
	Case (a)	0.44(0.04)	1.47(0.05)	0.76(0.05)	2.07(0.07)
	Case (b)	0.45(0.04)	1.49(0.05)	0.76(0.05)	2.07(0.07)
	Case (c)	0.47(0.04)	1.52(0.05)	0.76(0.05)	2.07(0.08)

S.10.2 Factor number selection

In this subsection, we report the accuracy of r selection. The values of Y_t are generated as described in Section 4.1 with $(p, T) \in \{50, 100, 150\} \times \{50, 100\}$ and $\{100, 150, 200\} \times \{300, 500\}$. The results for r selection are presented in Table S.2. Each simulation is replicated 100 times. We can correctly select r in all cases.

Table S.2: Correct Selection Rate of r for Different Values of p and T.

	$T = 50$	$T = 100$		$T = 300$	$T = 500$
$p = 50$	1	1	$p = 100$	1	1
$p = 100$	1	1	$p = 150$	1	1
$p = 150$	1	1	$p = 200$	1	1

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and K. Li (2012). Maximum likelihood estimation and inference for approximate factor models of high dimension. Technical report, University Library of Munich, Germany.

Bai, J. and Y. Liao (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of Econometrics 191*(1), 1–18.

Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis 7*(1), 1–46.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Latała, R. (2005). Some estimates of norms of random matrices. *Proceedings of the American Mathematical Society 133*(5), 1273–1282.

Ma, Z., Z. Ma, and H. Yuan (2020). Universal latent space model fitting for large networks with edge covariates. *The Journal of Machine Learning Research 21*(1), 86–152.

Yu, Y., T. Wang, and R. J. Samworth (2015). A useful variant of the davis–kahan theorem for statisticians. *Biometrika 102*(2), 315–323.

Zhang, X., S. Xue, and J. Zhu (2020). A flexible latent space model for multilayer networks. In

International Conference on Machine Learning, pp. 11288–11297. PMLR.