Supplementary materials of "Embracing the Blessing of Dimensionality in Factor Models"

Quefeng Li, Guang Cheng, Jianqin Fan, and Yuyan Wang

Additional Regularity Conditions

(iv) $\{\mathbf{u}_t, \mathbf{f}_t\}_{t\geq 1}$ are i.i.d. sub-Gaussian random variables over t.
(v) There exist constants c_1 and c_2 that $0 < c_1 \leq \lambda_{\min}(\mathbf{\Sigma}_u) \leq \lambda_{\max}(\mathbf{\Sigma}_u) \leq c_2 < \infty$, $\|\mathbf{\Sigma}_u\|_1 < c_2$ and $\min_{i\leq p, j\leq p} \mathrm{Var}(u_{it}u_{jt}) > c_1$;
(vi) There exists an $M > 0$ such that $||\mathbf{B}||_{\max} < M$;
(vii) There exists an $M > 0$ such that for any $s \leq T$ and $t \leq T$, $\mathrm{E}|p^{-1/2}(\mathbf{u}_s'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t - \mathrm{E}\mathbf{u}_s'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t)|^4 < M$ and $\mathrm{E}||p^{-1/2}\mathbf{B}'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t||^4 < M$;
(viii) For each $t \leq T$, $\mathrm{E}\|(pT)^{-1/2}\sum_{s=1}^T \mathbf{f}_s(\mathbf{u}_s'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t - \mathrm{E}(\mathbf{u}_s'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t))\|^2 = O(1)$;
(ix) For each $i \leq p$, $\mathrm{E}\|(pT)^{-1/2}\sum_{t=1}^T\sum_{j=1}^p \mathbf{d}_j(u_{jt}u_{it} - \mathrm{E}u_{jt}u_{it})\| = O(1)$, where $\mathbf{d}_j$ is the jth column of $\mathbf{B}'\mathbf{\Sigma}_u^{-1}$;
(x) For each $i \leq K$, $\mathrm{E}\|(pT)^{-1/2}\sum_{t=1}^T\sum_{j=1}^N \mathbf{d}_j u_{jt} f_{it}\| = O(1)$.

Condition (iv) is a standard assumption in order to establish the exponential type of concentration inequality for the elements in $\mathbf{u}_t$ and $\mathbf{f}_t$. Condition (v) requires $\mathbf{\Sigma}_u$ to be well-conditioned. In particular, we need a lower bound on the eigen-values of $\mathbf{\Sigma}_u$. This assumption guarantees that $\tilde{\mathbf{\Sigma}}_u$ is asymptotically non-singular so that $\tilde{\mathbf{\Sigma}}_u^{-1}$ will not perform badly in the weighted least-squares problem described in (6). These conditions were also assumed in Fan et al. (2013). Conditions (vii)-(x) are some moment conditions needed to establish the central limit theorem for the WPC estimator $\widehat{\mathbf{f}}_t$. They are standard in the factor model literature, e.g. Stock and Watson (2002) and Bai (2003).

Proofs of Results in Sections 2 and 4

Proof of Proposition 1. Let $\mathbf{g}_1 = \nabla_{\boldsymbol{\theta}_S} \log h(\mathbf{y}_S - \boldsymbol{\theta}_S, \mathbf{y}_{S^c} - \boldsymbol{\theta}_{S^c})$ and $\mathbf{g}_2 = \nabla_{\boldsymbol{\theta}_S} \log h_S(\mathbf{y}_S - \boldsymbol{\theta}_S)$, where h_S is the marginal density of $\mathbf{y}_S$. Firstly, we show that $\mathbf{g}_2 = \mathrm{E}(\mathbf{g}_1|\mathbf{y}_S)$. In fact, for any bounded function $\varphi(\mathbf{y}_S)$, by Fubini Theorem and condition (3),

$$
\begin{aligned}
\mathrm{E}(\mathbf{g}_1\varphi(\mathbf{y}_S)) &= -\iint (\nabla_{\mathbf{y}_S} \log h(\mathbf{y}_S - \boldsymbol{\theta}_S, \mathbf{y}_{S^c} - \boldsymbol{\theta}_{S^c}))h(\mathbf{y}_S - \boldsymbol{\theta}_S, \mathbf{y}_{S^c} - \boldsymbol{\theta}_{S^c})\varphi(\mathbf{y}_S)\mathrm{d}\mathbf{y}_S\mathrm{d}\mathbf{y}_{S^c} \\
&= -\iint (\nabla_{\mathbf{y}_S} h(\mathbf{y}_S - \boldsymbol{\theta}_S, \mathbf{y}_{S^c} - \boldsymbol{\theta}_{S^c}))\varphi(\mathbf{y}_S)\mathrm{d}\mathbf{y}_S\mathrm{d}\mathbf{y}_{S^c} \\
&= -\int \left(\nabla_{\mathbf{y}_S} \int h(\mathbf{y}_S - \boldsymbol{\theta}_S, \mathbf{y}_{S^c} - \boldsymbol{\theta}_{S^c})\mathrm{d}\mathbf{y}_{S^c}\right)\varphi(\mathbf{y}_S)\mathrm{d}\mathbf{y}_S
\end{aligned}
$$

$$
\begin{aligned}
&= -\int \nabla_{\mathbf{y}_S} h_S(\mathbf{y}_S - \boldsymbol{\theta}_S)\varphi(\mathbf{y}_S)\mathrm{d}\mathbf{y}_S \\
&= \int \left(\nabla_{\mathbf{y}_S} \log h_S(\mathbf{y}_S - \boldsymbol{\theta}_S)\right) h_S(\mathbf{y}_S - \boldsymbol{\theta}_S)\varphi(\mathbf{y}_S)\mathrm{d}\mathbf{y}_S \\
&= \mathrm{E}(\mathbf{g}_2\varphi(\mathbf{y}_S)).
\end{aligned}
$$

Then, by definition, $\mathbf{g}_2 = \mathrm{E}(\mathbf{g}_1|\mathbf{y}_S)$. Therefore,

$$
\begin{aligned}
\{I_p(\boldsymbol{\theta})\}_S &= \mathrm{E}(\mathbf{g}_1\mathbf{g}_1') = \mathrm{E}[(\mathbf{g}_2 + \mathbf{g}_1 - \mathbf{g}_2)(\mathbf{g}_2 + \mathbf{g}_1 - \mathbf{g}_2)'] \\
&= \mathrm{E}[\mathbf{g}_2\mathbf{g}_2'] + \mathrm{E}[\mathbf{g}_2(\mathbf{g}_1 - \mathbf{g}_2)'] + \mathrm{E}[(\mathbf{g}_1 - \mathbf{g}_2)\mathbf{g}_2'] + \mathrm{E}[(\mathbf{g}_1 - \mathbf{g}_2)(\mathbf{g}_1 - \mathbf{g}_2)'] \\
&= I_S(\boldsymbol{\theta}_S) + \mathrm{E}[(\mathbf{g}_1 - \mathbf{g}_2)(\mathbf{g}_1 - \mathbf{g}_2)'] \\
&\succeq I_S(\boldsymbol{\theta}_S),
\end{aligned}
$$

where the last equality follows from $\mathrm{E}[\mathbf{g}_2(\mathbf{g}_1 - \mathbf{g}_2)'] = \mathrm{E}[\mathrm{E}[\mathbf{g}_2(\mathbf{g}_1 - \mathbf{g}_2)'|\mathbf{y}_S]] = 0$, since $\mathbf{g}_2 = \mathrm{E}(\mathbf{g}_1|\mathbf{y}_S)$. □

Proof of Example 2. Without loss of generality, we assume $\boldsymbol{\theta} = \mathbf{0}$ so that the density of $\mathbf{y}$ is proportional to $g(\mathbf{y}'\boldsymbol{\Omega}\mathbf{y})$, where $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. Then,

$$
\begin{aligned}
|\nabla_{\mathbf{y}_S} h(\mathbf{y}_S, \mathbf{y}_{S^c})| &= 2\,|g'(\mathbf{y}'\boldsymbol{\Omega}\mathbf{y})(\boldsymbol{\Omega}\mathbf{y})_S| \leq 2\,|g'(\mathbf{y}'\boldsymbol{\Omega}\mathbf{y})|\,|\boldsymbol{\Omega}_S\mathbf{y}_S + \boldsymbol{\Omega}_{S,S^c}\mathbf{y}_{S^c}| \\
&\leq 2c\,|\boldsymbol{\Omega}_S\mathbf{y}_S + \boldsymbol{\Omega}_{S,S^c}\mathbf{y}_{S^c}|\,g(\mathbf{y}'\boldsymbol{\Omega}\mathbf{y}).
\end{aligned}
$$

Note that

$$
\begin{aligned}
\int \left(\int |\boldsymbol{\Omega}_S\mathbf{y}_S + \boldsymbol{\Omega}_{S,S^c}\mathbf{y}_{S^c}|\,g(\mathbf{y}'\boldsymbol{\Omega}\mathbf{y})d\mathbf{y}_{S^c} \right) d\mathbf{y}_S &\propto \mathrm{E}\left(|\boldsymbol{\Omega}_S\mathbf{y}_S + \boldsymbol{\Omega}_{S,S^c}\mathbf{y}_{S^c}|\right) \\
&\leq \mathrm{E}\left(|\boldsymbol{\Omega}_S\mathbf{y}_S| + |\boldsymbol{\Omega}_{S,S^c}\mathbf{y}_{S^c}|\right) \\
&< \infty.
\end{aligned}
$$

Therefore for a.e. any $\mathbf{y}_S$, $\int |\boldsymbol{\Omega}_S\mathbf{y}_S + \boldsymbol{\Omega}_{S,S^c}\mathbf{y}_{S^c}|\,g(\mathbf{y}'\boldsymbol{\Omega}\mathbf{y})$ is integrable. By Example 1.8 of Shao (2003), differentiation and integration are interchangeable, hence (3) holds. □

Proof of Proposition 2. For simplicity, let $\boldsymbol{\Omega} = I_p(\boldsymbol{\theta})$ and partition it as

$$
\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_S & \boldsymbol{\Omega}_{S,S^c} \\ \boldsymbol{\Omega}_{S^c,S} & \boldsymbol{\Omega}_{S^c} \end{pmatrix}.
$$

Then, the Fisher information $I(\mathbf{f})$ of $\mathbf{f}$ contained in all data is given by

$$
I(\mathbf{f}) = \mathbf{B}'\boldsymbol{\Omega}\mathbf{B} = \mathbf{B}_S'\boldsymbol{\Omega}_S\mathbf{B}_S + \mathbf{B}_{S^c}'\boldsymbol{\Omega}_{S^c,S}\mathbf{B}_S + \mathbf{B}_S'\boldsymbol{\Omega}_{S,S^c}\mathbf{B}_{S^c} + \mathbf{B}_{S^c}'\boldsymbol{\Omega}_{S^c}\mathbf{B}_{S^c}. \tag{A.1}
$$

If $\boldsymbol{\Omega}_{S,S^c} = \mathbf{0}$, we have

$$
\begin{aligned}
I(\mathbf{f}) &= \mathbf{B}_S'\boldsymbol{\Omega}_S\mathbf{B}_S + \mathbf{B}_{S^c}'\boldsymbol{\Omega}_{S^c}\mathbf{B}_{S^c} = \mathbf{B}_S'\{I_p(\boldsymbol{\theta})\}_S\mathbf{B}_S + \mathbf{B}_{S^c}'\boldsymbol{\Omega}_{S^c}\mathbf{B}_{S^c} \\
&\succeq \mathbf{B}_S'I_S(\boldsymbol{\theta}_S)\mathbf{B}_S + \mathbf{B}_{S^c}'\boldsymbol{\Omega}_{S^c}\mathbf{B}_{S^c} \succeq \mathbf{B}_S'I_S(\boldsymbol{\theta}_S)\mathbf{B}_S = I_S(\mathbf{f}),
\end{aligned}
$$

where the first inequality follows from Proposition 1 and the last inequality follows from that $\mathbf{B}_{S^c}'\boldsymbol{\Omega}_{S^c}\mathbf{B}_{S^c}$ is positive semi-definite. This completes the proof. □

Proof of Proposition 3. For any general $\mathbf{Q} \in \mathbb{R}^{L\times R}$, $\mathbf{B}_L \in \mathbb{R}^{L\times K}$, and $\mathbf{B}_R \in \mathbb{R}^{R\times K}$, we have

$$\mathrm{E}(\mathbf{B}_L'\mathbf{Q}\mathbf{B}_R) = \mathrm{E}\left[\sum_{l=1}^{L}\sum_{r=1}^{R} q_{l,r}\mathbf{b}_{L,l}\mathbf{b}_{R,r}'\right].$$

where $q_{l,r}$ is the (l,r)-th element of $\mathbf{Q}$, $\mathbf{b}_{L,l}'$ is the lth row of $\mathbf{B}_L$ and $\mathbf{b}_{R,r}'$ is the rth row of $\mathbf{B}_R$. Therefore,

$$\mathrm{E}(\mathbf{B}_{S^c}'\mathbf{\Omega}_{S^c,S}\mathbf{B}_S) = \mathrm{E}\left[\sum_{l\in S^C}\sum_{r\in S} \omega_{l,r}\mathbf{b}_{S^c,l}\mathbf{b}_{S,r}'\right],$$

where $\omega_{l,r}$ is the (l,r)-th element of $\mathbf{\Omega}$. By the i.i.d assumption, for $l \in S^C$ and $r \in S$, $\mathrm{E}(\mathbf{b}_{S^c,l}\mathbf{b}_{S,r}') = \mathrm{E}(\mathbf{b}_{S^c,l})\mathrm{E}(\mathbf{b}_{S,r}') = \mathbf{0}$. Hence, $\mathrm{E}(\mathbf{B}_{S^c}'\mathbf{\Omega}_{S^c,S}\mathbf{B}_S) = \mathbf{0}$. Similarly, it can be shown that $\mathrm{E}(\mathbf{B}_S'\mathbf{\Omega}_{S,S^c}\mathbf{B}_{S^c}) = \mathbf{0}$. By Proposition 1, $\mathbf{B}_S'\mathbf{\Omega}_S\mathbf{B}_S \succeq \mathbf{I}_S(\mathbf{f})$, which implies that $\mathrm{E}(\mathbf{B}_S'\mathbf{\Omega}_S\mathbf{B}_S) \succeq \mathrm{E}(\mathbf{I}_S(\mathbf{f}))$.

$$\mathrm{E}(\mathbf{B}_{S^c}'\mathbf{\Omega}_{S^c}\mathbf{B}_{S^c}) = \mathrm{E}\left[\sum_{l\in S^c}\sum_{r\in S^c} \omega_{l,r}\mathbf{b}_{L,l}\mathbf{b}_{R,r}'\right] = \mathrm{E}\left[\sum_{l\in S^c} \omega_{l,l}\mathbf{b}_{L,l}\mathbf{b}_{L,l}'\right] = \mathrm{tr}(\mathbf{\Omega}_{S^c})\mathrm{E}(\mathbf{b}\mathbf{b}') \succeq \mathbf{0}.$$

Using (A.1) and the above results, we have $\mathrm{E}[I(\mathbf{f})] \succeq \mathrm{E}[I_S(\mathbf{f})]$. □

Proof of Lemma 1. Since we assume all conditions hold for both s and p, we prove the result for p, i.e. $\max_{t\le T}\|\widehat{\mathbf{f}}_t^{(2)} - \mathbf{H}_2\mathbf{f}_t\| = O_P\left(T^{-1/2} + T^{1/4}/p^{-1/2}\right)$. The result for s can be proved similarly. For simplicity, we write $\widehat{\mathbf{f}}_t^{(2)}$ as $\widehat{\mathbf{f}}_t$ and $\mathbf{H}_2$ as $\mathbf{H}$.

By (A.1) of Bai and Liao (2013), $\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t$ has the following expansion,

$$\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t = \widehat{\mathbf{V}}^{-1}\left(\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i\mathbf{u}_i'\tilde{\mathbf{\Sigma}}_u^{-1}\mathbf{u}_t/p + \frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i\widehat{\eta}_{it} + \frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i\widehat{\theta}_{it}\right),$$

where $\widehat{\eta}_{it} = \mathbf{f}_i'\mathbf{B}'\tilde{\mathbf{\Sigma}}_u^{-1}\mathbf{u}_t/p$, $\widehat{\theta}_{it} = \mathbf{f}_t'\mathbf{B}'\tilde{\mathbf{\Sigma}}_u^{-1}\mathbf{u}_i/p$, and $\widehat{\mathbf{V}}$ is the diagonal matrix of the K largest eigenvalues of $\mathbf{Y}'\tilde{\mathbf{\Sigma}}_u^{-1}\mathbf{Y}/T$. Let $\eta_{it} = \mathbf{f}_i'\mathbf{B}'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t/p$ and $\theta_{it} = \mathbf{f}_t'\mathbf{B}'\mathbf{\Sigma}_u^{-1}\mathbf{u}_i/p$. Then, we have

$$\begin{aligned}\|\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t\| \le \|\widehat{\mathbf{V}}^{-1}\| \Bigg(&\left\|\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i\mathbf{u}_i'(\tilde{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1})\mathbf{u}_t/p\right\| + \left\|\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i(\mathbf{u}_i'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t - \mathrm{E}\mathbf{u}_i'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t)/p\right\| \\ &+ \left\|\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i\mathrm{E}(\mathbf{u}_i'\mathbf{\Sigma}_u^{-1}\mathbf{u}_t)/p\right\| + \left\|\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i(\widehat{\eta}_{it} - \eta_{it})\right\| + \left\|\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i\eta_{it}\right\| \\ &+ \left\|\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i(\widehat{\theta}_{it} - \theta_{it})\right\| + \left\|\frac{1}{T}\sum_{i=1}^{T}\widehat{\mathbf{f}}_i\theta_{it}\right\|\Bigg). \end{aligned} \tag{A.2}$$

Denote the jth summand inside the parenthesis as G_{jt}.

By Lemma A.2 of Bai and Liao (2013), $\|\widehat{\mathbf{V}}^{-1}\| = O_P(1)$. By Lemma A.6(iv) of Bai and Liao (2013),

$$\max_{t\le T} G_{1t} = O_P\left(\|\tilde{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1}\|\left\{\|\tilde{\mathbf{\Sigma}}_u^{-1} - \mathbf{\Sigma}_u^{-1}\| + 1/\sqrt{p} + \sqrt{(\log p)/T}\right\}\right).$$

By Proposition 4.1 of Bai and Liao (2013),

$$\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\| = o_P\left(\min\left\{T^{-1/4}, p^{-1/4}, \sqrt{T/(p\log p)}\right\}\right), \tag{A.3}$$

therefore, $\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\|\left(\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\| + 1/\sqrt{p} + \sqrt{(\log p)/T}\right) = o(T^{-1/2} + p^{-1/2})$. Hence,

$$\max_{t\leq T} G_{1t} = o_P\left(T^{-1/2} + p^{-1/2}\right).$$

By Lemma A.8(ii) of Bai and Liao (2013), $\max_{t\leq T} G_{2t} = O_P\left(T^{1/4}p^{-1/2}\right)$. By Lemma A.10(i) of Bai and Liao (2013), $\max_{t\leq T} G_{3t} = O_P\left(T^{-1/2}\right)$. By Lemma A.6(vi) of Bai and Liao (2013),

$$\max_{t\leq T} G_{4t} = O_P\left(\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\|\left\{\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\| + 1/\sqrt{p} + 1/\sqrt{T}\right\}\right) + o_P\left(1/\sqrt{p}\right) = o_P\left(1/\sqrt{p}\right).$$

By Lemma A.8(iii) of Bai and Liao (2013), $\max_{t\leq T} G_{5t} = O_P\left(T^{1/4}p^{-1/2}\right)$. By Lemma A.6(v) of Bai and Liao (2013) and (A.3),

$$\max_{t\leq T} G_{6t} = O_P\left(\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\|\left\{\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\| + 1/\sqrt{p} + \sqrt{(\log p)/T}\right\}\right) = o_P\left(1/\sqrt{p}\right).$$

By Lemma A.6(iii) of Bai and Liao (2013) and (A.3),

$$\max_{t\leq T} G_{7t} = O_P\left(\|\tilde{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\|/\sqrt{p} + 1/p + 1/\sqrt{pT}\right) = o_P\left(1/\sqrt{p}\right).$$

Then, by (A.2), we have

$$\max_{t\leq T}\|\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t\| = O_P\left(\frac{1}{\sqrt{T}} + \frac{T^{1/4}}{\sqrt{p}}\right).$$

□

Proof of Lemma 2. For Method 1, we have the following decomposition

$$\widehat{\mathbf{b}}_i^{(1)} - \mathbf{H}_1\mathbf{b}_i = \underbrace{\frac{1}{T}\sum_{t=1}^T \mathbf{H}_1\mathbf{f}_t u_{it}}_{I_1} + \underbrace{\frac{1}{T}\sum_{t=1}^T y_{it}(\widehat{\mathbf{f}}_t^{(1)} - \mathbf{H}_1\mathbf{f}_t)}_{I_2} + \underbrace{\mathbf{H}_1(\frac{1}{T}\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t' - \mathbf{I}_K)\mathbf{b}_i}_{I_3},$$

where $\mathbf{b}_i$ is the true factor loading of the ith subject as defined in (1).

For I_1, we have

$$\max_{i\leq s}\left\|\frac{1}{T}\sum_{t=1}^T \mathbf{H}_1\mathbf{f}_t u_{it}\right\| \leq \|\mathbf{H}_1\|\max_{i\leq s}\sqrt{\sum_{k=1}^K\left(\frac{1}{T}\sum_{t=1}^T f_{kt}u_{it}\right)^2}.$$

It follows from Lemma C.3(iii) of Fan et al. (2013) that, $\max_{i\leq s}\sqrt{\sum_{k=1}^K(\frac{1}{T}\sum_{t=1}^T f_{kt}u_{it})^2}$ $= O_P\left(\sqrt{(\log s)/T}\right)$. From Lemma A.2, $\|\mathbf{H}_1\| = O_P(1)$, therefore $I_1 = O_P\left(\sqrt{(\log s)/T}\right)$.

As for I_2, by conditions (v) and (vi),

$$\max_{i\le s}\mathrm{E}y_{it}^2=\max_{i\le s}\{\mathrm{E}(\mathbf{b}_i'\mathbf{f}_t)^2+\mathrm{E}u_{it}^2\}\le\max_{i\le s}\|\mathbf{b}_i\|^2+\max_{i\le s}\mathrm{Var}(u_{it})=O(1).$$

By condition (iv), y_{it}^2 is sub-exponential, therefore by the union bound and sub-exponential tail bound, $\max_{i\le s}\left|\frac{1}{T}\sum_{t=1}^T y_{it}^2-\mathrm{E}y_{it}^2\right|=O_P\left(\sqrt{(\log s)/T}\right)$. Then,

$$\max_{i\le s}\frac{1}{T}\sum_{t=1}^T y_{it}^2\le\max_{i\le s}\left|\frac{1}{T}\sum_{t=1}^T y_{it}^2-\mathrm{E}y_{it}^2\right|+\max_{i\le s}\mathrm{E}y_{it}^2=O_P\left(1\right). \tag{A.4}$$

By Cauchy-Schwartz inequality,

$$\begin{aligned}\max_{i\le s}\left\|\frac{1}{T}\sum_{t=1}^T y_{it}(\widehat{\mathbf{f}}_t^{(1)}-\mathbf{H}_1\mathbf{f}_t)\right\|&\le\max_{i\le s}\left(\frac{1}{T}\sum_{t=1}^T y_{it}^2\cdot\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{f}}_t^{(1)}-\mathbf{H}_1\mathbf{f}_t\|^2\right)^{1/2}\\&=O_P\left(\left(\frac{1}{T}\sum_{t=1}^T\|\widehat{\mathbf{f}}_t^{(1)}-\mathbf{H}_1\mathbf{f}_t\|^2\right)^{1/2}\right)\\&=O_P\left(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{s}}\right),\end{aligned}$$

where the last equality follows from Lemma A.5. So, $I_2=O_P\left(1/\sqrt{T}+1/\sqrt{s}\right)$.

Finally, it follows from Lemma C.3(i) of Fan et al. (2013) that $\|\frac{1}{T}\sum_{t=1}^T\mathbf{f}_t\mathbf{f}_t'-\mathbf{I}_K\|=O_P\left(T^{-1/2}\right)$. This together with $\|\mathbf{H}_1\|=O_P\left(1\right)$ and condition (vi) show that $I_3=O_P\left(T^{-1/2}\right)$. Hence,

$$\max_{i\le s}\|\widehat{\mathbf{b}}_i^{(1)}-\mathbf{H}_1\mathbf{b}_i\|=O_P\left(\frac{1}{\sqrt{s}}+\sqrt{\frac{\log s}{T}}\right).$$

Using the same arguments and the results of $\widehat{\mathbf{f}}_t^{(2)}$ in Lemma 1, we can show that

$$\max_{i\le s}\|\widehat{\mathbf{b}}_i^{(2)}-\mathbf{H}_2\mathbf{b}_i\|=O_P\left(\frac{1}{\sqrt{p}}+\sqrt{\frac{\log s}{T}}\right).$$

When the common factor $\mathbf{f}_t$ is known, for the oracle estimator of the loading matrix, we have

$$\begin{aligned}\max_{i\le s}\|\widehat{\mathbf{b}}_i^{o}-\mathbf{b}_i\|&\le\max_{i\le s}\left\|\frac{1}{T}\sum_{t=1}^T\mathbf{f}_t u_{it}\right\|+\left\|\frac{1}{T}\sum_{t=1}^T\mathbf{f}_t\mathbf{f}_t'-\mathbf{I}_K\right\|\max_{i\le s}\|\mathbf{b}_i\|\\&=O_P\left(\sqrt{\frac{\log s}{T}}+\frac{1}{\sqrt{T}}\right)\\&=O_P\left(\sqrt{\frac{\log s}{T}}\right).\end{aligned}$$

□

Proof of Lemma 3. By Theorem A.1 of Fan et al. (2013) (cited as Lemma A.7 in this document), it suffices to show

$$\max_{i\leq s}\frac{1}{T}\sum_{t=1}^{T}(u_{it}-\widehat{u}_{it}^{(1)})^2=O_P\left(\frac{1}{s}+\frac{\log s}{T}\right)\qquad\text{and}\qquad\max_{i,t}|u_{it}-\widehat{u}_{it}^{(1)}|=o_P\left(1\right).$$

For Method 1, we have

$$u_{it}-\widehat{u}_{it}^{(1)}=\mathbf{b}_i'\mathbf{H}_1'(\widehat{\mathbf{f}}_t^{(1)}-\mathbf{H}_1\mathbf{f}_t)+\{(\widehat{\mathbf{b}}_i^{(1)})'-\mathbf{b}_i'\mathbf{H}_1\}\widehat{\mathbf{f}}_t^{(1)}+\mathbf{b}_i'(\mathbf{H}_1'\mathbf{H}_1-\mathbf{I}_K)\mathbf{f}_t.$$

Using $(a+b+c)^2\leq 4a^2+4b^2+4c^2$, we have

$$\begin{aligned}\max_{i\leq s}\frac{1}{T}\sum_{t=1}^{T}(u_{it}-\widehat{u}_{it}^{(1)})^2\leq{}&4\max_{i\leq s}\|\mathbf{H}_1\mathbf{b}_i\|^2\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t^{(1)}-\mathbf{H}_1\mathbf{f}_t\|^2\\&+4\max_{i\leq s}\|\widehat{\mathbf{b}}_i^{(1)}-\mathbf{H}_1\mathbf{b}_i\|^2\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t^{(1)}\|^2\\&+4\|\mathbf{H}_1'\mathbf{H}_1-\mathbf{I}_K\|_F^2\max_{i\leq s}\|\mathbf{b}_i\|^2\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{f}_t\|^2.\end{aligned}$$

Since, $\max_i\|\mathbf{H}_1\mathbf{b}_i\|\leq\|\mathbf{H}_1\|\max_i\|\mathbf{b}_i\|=O_P\left(1\right)$, $\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t^{(1)}\|^2=O_P\left(1\right)$, and $\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{f}_t\|^2$ $=O_P\left(1\right)$, it follows from Lemma 1, 2, A.3 and A.5 that

$$\max_{i\leq s}\frac{1}{T}\sum_{t=1}^{T}(u_{it}-\widehat{u}_{it}^{(1)})^2=O_P\left(\frac{1}{s}+\frac{\log s}{T}\right).\tag{A.5}$$

On the other hand, by Lemma A.1,

$$\max_{i,t}|u_{it}-\widehat{u}_{it}^{(1)}|=\max_{i,t}|(\widehat{\mathbf{b}}_i^{(1)})'\widehat{\mathbf{f}}_i^{(1)}-\mathbf{b}_i'\mathbf{f}_t|=O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}}+\frac{T^{1/4}}{\sqrt{s}}\right)=o(1).$$

Then, the result follows from Theorem A.1 of Fan et al. (2013).

In analogous, a similar result can be proved for Method 2. For the oracle estimator, $\widehat{u}_{it}^{o}=y_{it}-(\widehat{\mathbf{b}}_i^{o})'\mathbf{f}_t$. Therefore,

$$\max_{i\leq s}\frac{1}{T}\sum_{t=1}^{T}(u_{it}-\widehat{u}_{it}^{o})^2\leq\max_{i\leq s}\|\widehat{\mathbf{b}}_i^{o}-\mathbf{b}_i\|^2\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{f}_t\|^2=O_P\left(\max_{i\leq s}\|\widehat{\mathbf{b}}_i^{o}-\mathbf{b}_i\|^2\right)=O_P\left(\frac{\log s}{T}\right).$$

$$\max_{i,t}|u_{it}-\widehat{u}_{it}^{o}|=\max_{i,t}|(\widehat{\mathbf{b}}_i^{o})'\mathbf{f}_t-\mathbf{b}_i'\mathbf{f}_t|=O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}}\right)=o_P(1).$$

It then follows from Theorem A.1 of Fan et al. (2013) that

$$\|\widehat{\mathbf{\Sigma}}_{u,S}^{o}-\mathbf{\Sigma}_{u,S}\|=O_P\left(m_s\sqrt{\frac{\log s}{T}}\right)=\|(\widehat{\mathbf{\Sigma}}_{u,S}^{o})^{-1}-\mathbf{\Sigma}_{u,S}^{-1}\|.$$

□

Proof of Theorem 1. (1) For Method 1, $\widehat{\mathbf{\Sigma}}_S^{(1)} = \widehat{\mathbf{B}}_1\widehat{\mathbf{B}}_1' + \widehat{\mathbf{\Sigma}}_{u,S}^{(1)}$. Therefore,

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}_S^{(1)} - \mathbf{\Sigma}_S\|_{\mathbf{\Sigma}_S}^2 &\le 2\left(\|\widehat{\mathbf{B}}_1\widehat{\mathbf{B}}_1' - \mathbf{B}_S\mathbf{B}_S'\|_{\mathbf{\Sigma}_S}^2 + \|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2\right)\\ &\le 2\big(\|\mathbf{B}_S(\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K)\mathbf{B}_S'\|_{\mathbf{\Sigma}_S}^2 + 2\|\mathbf{B}_S\mathbf{H}_1'\mathbf{C}_1'\|_{\mathbf{\Sigma}_S}^2 + \|\mathbf{C}_1\mathbf{C}_1'\|_{\mathbf{\Sigma}_S}^2\\ &\quad + \|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2\Big),\end{aligned}$$

where $\mathbf{C}_1 = \widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1'$. Then, it follows from Lemma A.4 that

$$\|\widehat{\mathbf{\Sigma}}_S^{(1)} - \mathbf{\Sigma}_S\|_{\mathbf{\Sigma}_S}^2 = O_P\left(\frac{1}{sT} + \frac{1}{s^2} + w_1^2 + sw_1^4 + m_s^2w_1^2\right) = O_P\left(sw_1^4 + m_s^2w_1^2\right).$$

Similarly, $\|\widehat{\mathbf{\Sigma}}_S^{(2)} - \mathbf{\Sigma}_S\|_{\mathbf{\Sigma}_S}^2 = O_P\left(sw_2^4 + m_s^2w_2^2\right)$.

In the oracle case, we have

$$\begin{aligned}\|\widehat{\mathbf{\Sigma}}_S^{o} - \mathbf{\Sigma}_S\|_{\mathbf{\Sigma}_S}^2 &\le 2\left(\|\widehat{\mathbf{B}}_o\widehat{\mathbf{B}}_o' - \mathbf{B}_S\mathbf{B}_S'\|_{\mathbf{\Sigma}_S}^2 + \|\widehat{\mathbf{\Sigma}}_{u,S}^{o} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2\right)\\ &\le 2\Bigg(\underbrace{\|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)(\widehat{\mathbf{B}}_o - \mathbf{B}_S)'\|_{\mathbf{\Sigma}_S}^2}_{I_1} + 2\underbrace{\|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\mathbf{B}_S'\|_{\mathbf{\Sigma}_S}^2}_{I_2} + \underbrace{\|\widehat{\mathbf{\Sigma}}_{u,S}^{o} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2}_{I_3}\Bigg).\end{aligned}$$

Since all eigenvalues of $\mathbf{\Sigma}_S$ are bounded away from zero, for any matrix $\mathbf{A} \in \mathbb{R}^{s\times s}$, $\|\mathbf{A}\|_{\mathbf{\Sigma}_S}^2 = s^{-1}\|\mathbf{\Sigma}^{-1/2}\mathbf{A}\mathbf{\Sigma}^{-1/2}\|_F^2 = O_P\left(s^{-1}\|\mathbf{A}\|_F^2\right)$. Then, by Lemma 2, we have

$$I_1 = O_P\left(s^{-1}\|\widehat{\mathbf{B}}_o - \mathbf{B}_S\|_F^4\right) = O_P\left(sw_o^4\right),$$

where the last equality follows that $\|\widehat{\mathbf{B}}_o - \mathbf{B}_S\|_F^2 \le s(\max_{i\le s}\|\widehat{\mathbf{b}}_i^o - \mathbf{b}_i\|)^2 = O_P\left(sw_o^2\right)$. For I_2, we have

$$\begin{aligned}I_2 &= s^{-1}\mathrm{tr}((\widehat{\mathbf{B}}_o - \mathbf{B}_S)'\mathbf{\Sigma}_S^{-1}(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\mathbf{B}_S'\mathbf{\Sigma}_S^{-1}\mathbf{B}_S)\\ &\le s^{-1}\|\mathbf{\Sigma}_S^{-1}\|\|\widehat{\mathbf{B}}_o - \mathbf{B}_S\|_F^2\|\mathbf{B}_S'\mathbf{\Sigma}_S^{-1}\mathbf{B}_S\|\\ &= O_P\left(w_o^2\right).\end{aligned}$$

For I_3, Lemma 3 implies that

$$I_3 = O_P\left(s^{-1}\|\widehat{\mathbf{\Sigma}}_{u,S}^{o} - \mathbf{\Sigma}_{u,S}\|_F^2\right) = O_P\left(\|\widehat{\mathbf{\Sigma}}_{u,S}^{o} - \mathbf{\Sigma}_{u,S}\|^2\right) = O_P\left(m_s^2w_o^2\right).$$

Therefore, $\|\widehat{\mathbf{\Sigma}}_{u,S}^{o} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2 = O_P\left(sw_o^4 + m_s^2w_o^2\right)$.

(2) For Method 1,

$$\|\widehat{\mathbf{\Sigma}}_S^{(1)} - \mathbf{\Sigma}_S\|_{\max} \le \underbrace{\|\widehat{\mathbf{B}}_1\widehat{\mathbf{B}}_1' - \mathbf{B}_S\mathbf{B}_S'\|_{\max}}_{I_1} + \underbrace{\|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)} - \mathbf{\Sigma}_{u,S}\|_{\max}}_{I_2}.$$

For I_1, we have

$$I_1 = \max_{ij}|(\widehat{\mathbf{b}}_i^{(1)})'\widehat{\mathbf{b}}_j^{(1)} - \mathbf{b}_i'\mathbf{b}_j|$$

$$
\begin{aligned}
&\leq \max_{ij}\left(|(\widehat{\mathbf{b}}_i^{(1)}-\mathbf{H}_1\mathbf{b}_i)'(\widehat{\mathbf{b}}_j^{(1)}-\mathbf{H}_1\mathbf{b}_j)|+2|\mathbf{b}_i'\mathbf{H}_1'(\widehat{\mathbf{b}}_j^{(1)}-\mathbf{H}_1\mathbf{b}_j)|+|\mathbf{b}_i'(\mathbf{H}_1\mathbf{H}_1'-\mathbf{I}_K)\mathbf{b}_j|\right)\\
&\leq \left(\max_i\|\widehat{\mathbf{b}}_i^{(1)}-\mathbf{H}_1\mathbf{b}_i\|\right)^2+2\max_{ij}\|\widehat{\mathbf{b}}_i^{(1)}-\mathbf{H}_1\mathbf{b}_i\|\|\mathbf{H}_1\mathbf{b}_j\|+\|\mathbf{H}_1\mathbf{H}_1'-\mathbf{I}_K\|\left(\max_i\|\mathbf{b}_i\|\right)^2\\
&= O_P\left(w_1\right),
\end{aligned}
$$

where the last identity follows from Lemmas 2 and A.3.

For I_2, let $\sigma_{u,ij}$ be the (i,j)-th entry of $\mathbf{\Sigma}_{u,S}$ and $\widehat{\sigma}_{u,ij}=\frac{1}{T}\sum_{t=1}^T\widehat{u}_{it}\widehat{u}_{jt}$, where $\widehat{u}_{it}$ are the estimator of u_{it} from Method 1 as described in Section 4. Then,

$$
\begin{aligned}
&\max_{ij}|\widehat{\sigma}_{u,ij}-\sigma_{u,ij}|\\
&=\max_{ij}\left|\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it}\widehat{u}_{jt}-u_{it}u_{jt})\right|+\max_{ij}\left|\frac{1}{T}\sum_{i=1}^T u_{it}u_{jt}-\mathrm{E}(u_{it}u_{jt})\right|\\
&\leq\max_{ij}\left|\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it}-u_{it})(\widehat{u}_{jt}-u_{jt})\right|+2\max_{ij}\left|\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it}-u_{it})u_{jt}\right|+\max_{ij}\left|\frac{1}{T}\sum_{i=1}^T u_{it}u_{jt}-\mathrm{E}(u_{it}u_{jt})\right|\\
&\leq\max_{ij}\left(\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it}-u_{it})^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{jt}-u_{jt})^2\right)^{1/2}+2\max_{ij}\left(\frac{1}{T}\sum_{t=1}^T(\widehat{u}_{it}-u_{it})^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^T u_{jt}^2\right)^{1/2}\\
&\quad+\max_{ij}\left|\frac{1}{T}\sum_{i=1}^T u_{it}u_{jt}-\mathrm{E}(u_{it}u_{jt})\right|\\
&=O_P\left(w_1^2\right)+O_P\left(w_1\right)+O_P\left(\sqrt{(\log s)/T}\right),
\end{aligned}
$$

where the last equality follows from (A.5), Lemma C.3 (ii) of Fan et al. (2013) and

$$
\max_{j\leq s}\frac{1}{T}\sum_{t=1}^T u_{jt}^2=O_P\left(1\right)
$$

as similarly shown in (A.4). Hence, $\max_{ij}|\widehat{\sigma}_{u,ij}-\sigma_{u,ij}|=O_P\left(w_1\right)$. After the thresholding,

$$
\begin{aligned}
\max_{ij}|s_{ij}(\widehat{\sigma}_{u,ij})-\sigma_{u,ij}|&\leq\max_{ij}|s_{ij}(\widehat{\sigma}_{u,ij})-\widehat{\sigma}_{u,ij}|+|\widehat{\sigma}_{u,ij}-\sigma_{u,ij}|\\
&\leq\max_{ij}|s_{ij}(\widehat{\sigma}_{u,ij})-\widehat{\sigma}_{u,ij}|+O_P\left(w_1\right)\\
&=O_P\left(w_1\right).
\end{aligned}
$$

where $s_{ij}(\cdot)$ is the hard thresholding at the level defined in step ii. of Method 1. Hence, $\|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)}-\mathbf{\Sigma}_{u,S}\|_{\max}=O_P\left(w_1\right)$. Similarly, $\|\widehat{\mathbf{\Sigma}}_{u,S}^{(2)}-\mathbf{\Sigma}_{u,S}\|_{\max}=O_P\left(w_2\right)$. For the oracle estimator,

$$
\begin{aligned}
\|\widehat{\mathbf{B}}_o\widehat{\mathbf{B}}_o'-\mathbf{B}\mathbf{B}'\|_{\max}&=\max_{ij}\left(|(\widehat{\mathbf{b}}_i^o-\mathbf{b}_i)'(\widehat{\mathbf{b}}_i-\mathbf{b}_i)|+2|(\widehat{\mathbf{b}}_i^o-\mathbf{b}_i)'\mathbf{b}_j|\right)\\
&\leq\left(\max_i\|\widehat{\mathbf{b}}_i^o-\mathbf{H}_1\mathbf{b}_i\|\right)^2+2\max_{ij}\|\widehat{\mathbf{b}}_i^o-\mathbf{b}_i\|\|\mathbf{b}_j\|
\end{aligned}
$$

$$= O_P\left(w_o\right),$$

where the last equality follows from condition (vi) and Lemma 2. Using similar arguments as in the above, $\max_{ij}|\widehat{\sigma}^o_{u,ij} - \sigma_{u,ij}| = O_P(w_0)$. Hence, $\|\widehat{\boldsymbol{\Sigma}}^o_{u,S} - \boldsymbol{\Sigma}_{u,S}\|_{\max} = O_P\left(w_o\right)$.

(3) For Method 1, let $\tilde{\boldsymbol{\Sigma}}_S = \mathbf{B}_S\mathbf{H}_1'\mathbf{H}_1\mathbf{B}_S' + \boldsymbol{\Sigma}_{u,S}$. We have

$$\|(\widehat{\boldsymbol{\Sigma}}^{(1)}_S)^{-1} - \boldsymbol{\Sigma}_S^{-1}\| \le \|(\widehat{\boldsymbol{\Sigma}}^{(1)}_S)^{-1} - \tilde{\boldsymbol{\Sigma}}_S^{-1}\| + \|\tilde{\boldsymbol{\Sigma}}_S^{-1} - \boldsymbol{\Sigma}_S^{-1}\|.$$

Since $\widehat{\boldsymbol{\Sigma}}^{(1)}_S = \widehat{\mathbf{B}}_1\widehat{\mathbf{B}}_1' + \widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S}$, by Sherman-Morrison-Woodbury formula,

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_S^{-1} &= \boldsymbol{\Sigma}_{u,S}^{-1} + \boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1'\mathbf{G}^{-1}\mathbf{H}_1\mathbf{B}_S\boldsymbol{\Sigma}_{u,S}^{-1}, \\
(\widehat{\boldsymbol{\Sigma}}^{(1)}_S)^{-1} &= (\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1} + (\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1}\widehat{\mathbf{B}}_1\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1},
\end{aligned}$$

where $\mathbf{G} = \mathbf{I}_K + \mathbf{H}_1\mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1'$ and $\widehat{\mathbf{G}} = \mathbf{I}_K + \widehat{\mathbf{B}}_1'(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1}\widehat{\mathbf{B}}_1$. Therefore, $\|(\widehat{\boldsymbol{\Sigma}}^{(1)}_S)^{-1} - \tilde{\boldsymbol{\Sigma}}_S^{-1}\| \le \sum_{i=1}^6 I_i$, where

$$\begin{aligned}
I_1 &= \|(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1} - \boldsymbol{\Sigma}_{u,S}^{-1}\|, \\
I_2 &= \|\{(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1} - \boldsymbol{\Sigma}_{u,S}^{-1}\}\widehat{\mathbf{B}}_1\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1}\|, \\
I_3 &= \|\{(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1} - \boldsymbol{\Sigma}_{u,S}^{-1}\}\widehat{\mathbf{B}}_1\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\boldsymbol{\Sigma}_{u,S}^{-1}\|, \\
I_4 &= \|\boldsymbol{\Sigma}_{u,S}^{-1}(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\boldsymbol{\Sigma}_{u,S}^{-1}\|, \\
I_5 &= \|\boldsymbol{\Sigma}_{u,S}^{-1}(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')\widehat{\mathbf{G}}^{-1}\mathbf{H}_1\mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\|, \\
I_6 &= \|\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1'\{\widehat{\mathbf{G}}^{-1} - \mathbf{G}^{-1}\}\mathbf{H}_1\mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\|.
\end{aligned}$$

From Lemma 3, $I_1 = O_P\left(m_s w_1\right)$. For I_2, we have

$$I_2 \le \|(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1} - \boldsymbol{\Sigma}_{u,S}^{-1}\|\|\widehat{\mathbf{B}}_1\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\|\|(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1}\|.$$

By Lemma 3 and condition (v), $\|(\widehat{\boldsymbol{\Sigma}}^{(1)}_{u,S})^{-1}\| = O_P\left(1\right)$. Lemma A.6(ii) implies that $\|\widehat{\mathbf{G}}^{-1}\| = O_P\left(s^{-1}\right)$. Therefore, $\|\widehat{\mathbf{B}}_1\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\| = O_P\left(1\right)$ and $I_2 = O_P\left(m_s w_1\right)$. Similarly, $I_3 = O_P\left(m_s w_1\right)$. For I_4, condition (v) implies that $\|\boldsymbol{\Sigma}_{u,S}^{-1}\| = O(1)$. Next, $\|(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\|$ is bounded by

$$\|(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\| \le \|(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')\widehat{\mathbf{G}}^{-1}(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')'\|^{1/2}\|\widehat{\mathbf{B}}_1\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\|^{1/2}.$$

Since $\|\widehat{\mathbf{G}}^{-1}\| = O_P(s^{-1})$ by Lemma A.6(ii) and $\|\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1'\|_F^2 = O_P\left(s w_1^2\right)$ by Lemma A.4(i), we have $\|(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')\widehat{\mathbf{G}}^{-1}(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')'\| = O_P\left(w_1^2\right)$. This together with $\|\widehat{\mathbf{B}}_1\widehat{\mathbf{G}}^{-1}\widehat{\mathbf{B}}_1'\| = O_P\left(1\right)$ imply that $I_4 = O_P\left(w_1\right)$. Similarly, $I_5 = O_P\left(w_1\right)$. For I_6, we have

$$I_6 \le \|\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1'\mathbf{H}_1\mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\|\|\widehat{\mathbf{G}}^{-1} - \mathbf{G}^{-1}\|.$$

Condition (ii), (v) and $\|\mathbf{H}_1\| = O_P(1)$ imply that $\|\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1'\mathbf{H}_1\mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\| = O_P(s)$. Next, we bound $\|\widehat{\mathbf{G}}^{-1} - \mathbf{G}^{-1}\|$. Note that,

$$\begin{aligned}\|\widehat{\mathbf{G}}^{-1} - \mathbf{G}^{-1}\| &= \|\mathbf{G}^{-1}(\widehat{\mathbf{G}} - \mathbf{G})\widehat{\mathbf{G}}^{-1}\| = O_P\left(s^{-2}\|\widehat{\mathbf{B}}_1'(\widehat{\mathbf{\Sigma}}_{u,S}^{(1)})^{-1}\widehat{\mathbf{B}}_1 - (\mathbf{B}_S\mathbf{H}_1')'\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1'\|\right)\\ &= O_P\left(s^{-1}m_s w_1\right),\end{aligned}$$

because by Lemma A.6 (i) and (ii), $\|\mathbf{G}^{-1}\| = O(s^{-1})$, $\|\widehat{\mathbf{G}}^{-1}\| = O_P(s^{-1})$, and

$$\begin{aligned}&\|\widehat{\mathbf{B}}_1'(\widehat{\mathbf{\Sigma}}_{u,S}^{(1)})^{-1}\widehat{\mathbf{B}}_1 - (\mathbf{B}_S\mathbf{H}_1')'\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1'\|\\ &\leq \|(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')'(\widehat{\mathbf{\Sigma}}_{u,S}^{(1)})^{-1}(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')\| + 2\|(\widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1')(\widehat{\mathbf{\Sigma}}_{u,S}^{(1)})^{-1}\mathbf{B}_S\mathbf{H}_1'\|\\ &\quad + \|(\mathbf{B}_S\mathbf{H}_1')'\{(\widehat{\mathbf{\Sigma}}_{u,S}^{(1)})^{-1} - \mathbf{\Sigma}_{u,S}^{-1}\}\mathbf{B}_S\mathbf{H}_1'\|\\ &= O_P\left(sw_1^2\right) + O_P(sw_1) + O_P(sm_s w_1)\\ &= O_P(sm_s w_1). \end{aligned} \tag{A.6}$$

Therefore, $I_6 = O_P(m_s w_1)$. Summing the six terms, we have $\|(\widehat{\mathbf{\Sigma}}_{u,S}^{(1)})^{-1} - \tilde{\mathbf{\Sigma}}_S^{-1}\| = O_P(m_s w_1)$. Next, we bound $\|\tilde{\mathbf{\Sigma}}_S^{-1} - \mathbf{\Sigma}_S^{-1}\|$.

By using Sherman-Morrison-Woodbury formula again,

$$\begin{aligned}\|\tilde{\mathbf{\Sigma}}_S^{-1} - \mathbf{\Sigma}_S^{-1}\| &= \left\|\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S\{[(\mathbf{H}_1'\mathbf{H}_1)^{-1} + \mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S]^{-1} - [\mathbf{I}_K + \mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S]^{-1}\}\mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\right\|\\ &= O(s)\left\|[(\mathbf{H}_1'\mathbf{H}_1)^{-1} + \mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S]^{-1} - [\mathbf{I}_K + \mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S]^{-1}\right\|\\ &= O_P\left(s^{-1}\right)\|(\mathbf{H}_1'\mathbf{H}_1)^{-1} - \mathbf{I}_K\|\\ &= o_P(m_s w_1).\end{aligned}$$

Therefore, $\|(\widehat{\mathbf{\Sigma}}_{u,S}^{(1)})^{-1} - \mathbf{\Sigma}_S^{-1}\| = O_P(m_s w_1)$. A similar result can be shown that $\|(\widehat{\mathbf{\Sigma}}_{u,S}^{(2)})^{-1} - \mathbf{\Sigma}_S^{-1}\| = O_P(m_s w_2)$.

For the oracle estimator, by Sherman-Morrison-Woodbury formula, $\|(\widehat{\mathbf{\Sigma}}_S^{o})^{-1} - \mathbf{\Sigma}_S^{-1}\| \leq \sum_{i=1}^6 I_i$, where

$$\begin{aligned}I_1 &= \|(\widehat{\mathbf{\Sigma}}_{u,S}^{o})^{-1} - \mathbf{\Sigma}_{u,S}^{-1}\|,\\ I_2 &= \|\{(\widehat{\mathbf{\Sigma}}_{u,S}^{o})^{-1} - \mathbf{\Sigma}_{u,S}^{-1}\}\widehat{\mathbf{B}}_o\widehat{\mathbf{J}}^{-1}\widehat{\mathbf{B}}_o'(\widehat{\mathbf{\Sigma}}_{u,S}^{o})^{-1}\|,\\ I_3 &= \|\{(\widehat{\mathbf{\Sigma}}_{u,S}^{o})^{-1} - \mathbf{\Sigma}_{u,S}^{-1}\}\widehat{\mathbf{B}}_o\widehat{\mathbf{J}}^{-1}\widehat{\mathbf{B}}_o'\mathbf{\Sigma}_{u,S}^{-1}\|,\\ I_4 &= \|\mathbf{\Sigma}_{u,S}^{-1}(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\widehat{\mathbf{J}}^{-1}\widehat{\mathbf{B}}_o'\mathbf{\Sigma}_{u,S}^{-1}\|,\\ I_5 &= \|\mathbf{\Sigma}_{u,S}^{-1}(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\widehat{\mathbf{J}}^{-1}\mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\|,\\ I_6 &= \|\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S\{\widehat{\mathbf{J}}^{-1} - \mathbf{J}^{-1}\}\mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\|,\end{aligned}$$

that $\widehat{\mathbf{J}} = \mathbf{I}_K + \widehat{\mathbf{B}}_o'(\widehat{\mathbf{\Sigma}}_{u,S}^{o})^{-1}\widehat{\mathbf{B}}_o$ and $\mathbf{J} = \mathbf{I}_K + \mathbf{B}_S'\mathbf{\Sigma}_{u,S}^{-1}\mathbf{B}_S$.

By Lemma 3, $I_1 = O_P(m_s w_o)$. For I_2, Lemma A.6(ii) implies that $\|\widehat{\mathbf{J}}^{-1}\| = O_P(s^{-1})$. This together with condition (ii) imply that $\|\widehat{\mathbf{B}}_o\widehat{\mathbf{J}}^{-1}\widehat{\mathbf{B}}_o'\| = O_P(1)$. Moreover, it follows from Lemma 3 and condition (v) that $\|(\widehat{\boldsymbol{\Sigma}}^o_{u,S})^{-1}\| = O_P(1)$. Therefore,

$$I_2 \leq \|(\widehat{\boldsymbol{\Sigma}}^o_{u,S})^{-1} - \boldsymbol{\Sigma}_{u,S}^{-1}\|\|\widehat{\mathbf{B}}_o\widehat{\mathbf{J}}^{-1}\widehat{\mathbf{B}}_o'\|\|(\widehat{\boldsymbol{\Sigma}}^o_{u,S})^{-1}\| = O_P(m_s w_o).$$

Similarly, $I_3 = O_P(m_s w_o)$. For I_4, we have $I_4 \leq \|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\widehat{\mathbf{J}}^{-1}\mathbf{B}_S'\|\|\boldsymbol{\Sigma}_{u,S}^{-1}\|^2$. We bound $\|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\widehat{\mathbf{J}}^{-1}\mathbf{B}_S'\|$ by

$$\|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\widehat{\mathbf{J}}^{-1}\mathbf{B}_S'\| \leq \|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\widehat{\mathbf{J}}^{-1}(\widehat{\mathbf{B}}_o - \mathbf{B}_S)'\|^{1/2}\|\mathbf{B}_S\widehat{\mathbf{J}}^{-1}\mathbf{B}_S'\|^{1/2}.$$

Since $\|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)(\widehat{\mathbf{B}}_o - \mathbf{B}_S)'\| \leq \|\widehat{\mathbf{B}}_o - \mathbf{B}_S\|_F^2 \leq s(\max_s\|\widehat{\mathbf{b}}_i^o - \mathbf{b}_i\|)^2 = O_P(sw_o^2)$. This together with $\|\widehat{\mathbf{J}}^{-1}\| = O_P(s^{-1})$ and $\|\widehat{\mathbf{B}}_o\widehat{\mathbf{J}}^{-1}\widehat{\mathbf{B}}_o\| = O_P(1)$ imply that $I_4 = O_P(w_o)$. Similarly, $I_5 = O_P(w_o)$. For I_6, we have $I_6 \leq \|\widehat{\mathbf{J}}^{-1} - \mathbf{J}^{-1}\|\|\boldsymbol{\Sigma}_{u,S}^{-1}\|^2\|\mathbf{B}_S\mathbf{B}_S'\|$. By conditions (ii) and (iv), we have $\|\boldsymbol{\Sigma}_{u,S}^{-1}\| = O(1)$ and $\|\mathbf{B}_S\mathbf{B}_S'\| = O(s)$. As for $\|\widehat{\mathbf{J}}^{-1} - \mathbf{J}^{-1}\|$, we have

$$\|\widehat{\mathbf{J}}^{-1} - \mathbf{J}^{-1}\| = \|\widehat{\mathbf{J}}^{-1}(\widehat{\mathbf{J}} - \mathbf{J})\mathbf{J}^{-1}\| = O_P\left(s^{-2}\|\mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S - \widehat{\mathbf{B}}_o'\widehat{\boldsymbol{\Sigma}}_{u,S}^{-1}\widehat{\mathbf{B}}_o\|\right) = O_P\left(s^{-1}m_s w_o\right),$$

where the last equation follows from that

$$\begin{aligned}\|\widehat{\mathbf{B}}_o'\widehat{\boldsymbol{\Sigma}}_{u,S}^{-1}\widehat{\mathbf{B}}_o - \mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\| &\leq \|(\widehat{\mathbf{B}}_o - \mathbf{B}_S)'\widehat{\boldsymbol{\Sigma}}_{u,S}^{-1}(\widehat{\mathbf{B}}_o - \mathbf{B}_S)\| + 2\|(\widehat{\mathbf{B}_o} - \mathbf{B}_S)'\widehat{\boldsymbol{\Sigma}}_{u,S}^{-1}\mathbf{B}_S\| \\ &\quad + \|\mathbf{B}_S'\{(\widehat{\boldsymbol{\Sigma}}^o_{u,S})^{-1} - \boldsymbol{\Sigma}_{u,S}^{-1}\}\mathbf{B}_S\| \\ &= O_P\left(sw_o^2\right) + O_P(sw_o) + O_P(sm_s w_o) \\ &= O_P(sm_s w_o).\end{aligned}$$

Therefore, $I_6 = O_P(m_s w_o)$. After summing up, $\|(\widehat{\boldsymbol{\Sigma}}^o_S)^{-1} - \boldsymbol{\Sigma}_S^{-1}\| = O_P(m_s w_o)$. □

Convergence Rates of $\bar{\boldsymbol{\Sigma}}_S$ in Section 5

Let $\bar{\mathbf{H}} = M^{-1}\sum_{m=1}^M \mathbf{H}_{[m]}$, where $\mathbf{H}_{[m]} = \widehat{\mathbf{V}}_m^{-1}\widehat{\mathbf{F}}_m'\mathbf{F}_m\mathbf{B}_m'\tilde{\boldsymbol{\Sigma}}_{u,m}^{-1}\mathbf{B}_m/T$, $\widehat{\mathbf{V}}_m$ is the diagonal matrix of the K largest eigenvalues of $\mathbf{Y}_m'\tilde{\boldsymbol{\Sigma}}_{u,m}^{-1}\mathbf{Y}_m/T$, $\mathbf{B}_m$ and $\mathbf{F}_m$ are the loadings and the factors in the mth group.

According to the proof of Theorem 1, the key is to show that $\max_{1\leq t\leq T}\|\bar{\mathbf{f}}_t - \bar{\mathbf{H}}\mathbf{f}_t\|$ has the same rate as $\max_{1\leq t\leq T}\|\widehat{\mathbf{f}}_t^{(2)} - \mathbf{H}_2\mathbf{f}_t\|$ and $\max_{i\leq s}\|\bar{\mathbf{b}}_i - \bar{\mathbf{H}}\mathbf{b}_i\|$ has the same rate as $\max_{1\leq i\leq s}\|\widehat{\mathbf{b}}_i^{(2)} - \mathbf{H}_2\mathbf{b}_i\|$.

To give the rate of $\max_{1\leq t\leq T}\|\bar{\mathbf{f}}_t - \bar{\mathbf{H}}\mathbf{f}_t\|$, since M is fixed, p/M is in the same order as p. Then, it follows from Lemma 1 that for any $1 \leq m \leq M$, $\max_{1\leq t\leq T}\|\widehat{\mathbf{f}}_{m,t} - \mathbf{H}_{[m]}\mathbf{f}_t\| = O_P(a_{p,T})$, where $a_{p,T} = T^{-1/2} + T^{1/4}p^{-1/2}$. By definition, there exists a positive constant $C_{m,\epsilon}$ such that

$$P\left(\max_{1\leq t\leq T}\|\widehat{\mathbf{f}}_{m,t} - \mathbf{H}_{[m]}\mathbf{f}_t\| > C_{m,\epsilon}a_{p,T}\right) \leq \epsilon/M.$$

Let $C = \max_{1\le m\le M} C_{m,\epsilon}$. We have

$$P\left(\max_{1\le t\le T}\|\bar{\mathbf{f}}_t - \bar{\mathbf{H}}\mathbf{f}_t\| > Ca_{p,T}\right) = P\left(\max_{1\le t\le T}\left\|\frac{1}{M}\sum_{m=1}^{M}(\widehat{\mathbf{f}}_{m,t} - \mathbf{H}_{[m]}\mathbf{f}_t)\right\| > Ca_{p,T}\right)$$
$$\le \sum_{m=1}^{M} P\left(\max_{1\le t\le T}\|\widehat{\mathbf{f}}_{m,t} - \mathbf{H}_{[m]}\mathbf{f}_t\| > Ca_{p,T}\right)$$
$$\le \epsilon.$$

By definition, $\max_{1\le t\le T}\|\bar{\mathbf{f}}_t - \bar{\mathbf{H}}\mathbf{f}_t\| = O_P(a_{p,T})$, which is the same as $\max_{1\le t\le T}\|\widehat{\mathbf{f}}_t^{(2)} - \mathbf{H}_2\mathbf{f}_t\|$ shown in Lemma 1.

Next, we show that $\max_{i\le s}\|\bar{\mathbf{b}}_i - \bar{\mathbf{H}}\mathbf{b}_i\| = O_P(w_2)$. For any $1 \le m \le M$, similarly as in Lemma A.2, we have $\|\mathbf{H}_{[m]}\| = O_P(1)$. By the same union bound argument, we have $\|\bar{\mathbf{H}}\| = O_P(1)$. Then, it follows from the same proof of Lemma 2 that $\max_{i\le s}\|\bar{\mathbf{b}}_i - \bar{\mathbf{H}}\mathbf{b}_i\| = O_P(w_2)$.

As M is fixed, the results in Lemma 3 and Theorem 1 for each individual group hold. Repeatedly using the above union bound argument, $\bar{\mathbf{\Sigma}}_S$ is shown to have the same convergence rate as $\widehat{\mathbf{\Sigma}}_S^{(2)}$.

Additional Lemmas

Lemma A.1. *Under conditions of Lemma 1, it holds that*

$$\max_{i\le s,t\le T}\|(\widehat{\mathbf{b}}_i^{(1)})'\widehat{\mathbf{f}}_t^{(1)} - \mathbf{b}_i'\mathbf{f}_t\| = O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}} + \frac{T^{1/4}}{\sqrt{s}}\right),$$
$$\max_{i\le s,t\le T}\|(\widehat{\mathbf{b}}_i^{(2)})'\widehat{\mathbf{f}}_t^{(2)} - \mathbf{b}_i'\mathbf{f}_t\| = O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}} + \frac{T^{1/4}}{\sqrt{p}}\right),$$
$$\max_{i\le s,t\le T}\|(\widehat{\mathbf{b}}_i^{o})'\mathbf{f}_t - \mathbf{b}_i'\mathbf{f}_t\| = O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}}\right).$$

Proof of Lemma A.1. Under condition (i), it follows from the union bound argument that
$$\max_{t\le T}\|\mathbf{f}_t\| = O_P\left(\sqrt{\log T}\right).$$

Then, for Method 1, it follows from Lemmas 1, 2, A.2, and condition (vi) that, uniformly in i and t,

$$\|(\widehat{\mathbf{b}}_i^{(1)})'\widehat{\mathbf{f}}_t^{(1)} - \mathbf{b}_i'\mathbf{f}_t\| \le \|\widehat{\mathbf{b}}_i^{(1)} - \mathbf{H}_1\mathbf{b}_i\|\|\widehat{\mathbf{f}}_t^{(1)} - \mathbf{H}_1\mathbf{f}_t\| + \|\mathbf{H}_1\mathbf{b}_i\|\|\widehat{\mathbf{f}}_t^{(1)} - \mathbf{H}_1\mathbf{f}_t\|$$
$$+ \|\widehat{\mathbf{b}}_i^{(1)} - \mathbf{H}_1\mathbf{b}_i\|\|\mathbf{H}_1\mathbf{f}_t\| + \|\mathbf{b}_i\|\|\mathbf{f}_t\|\|\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K\|_F$$
$$= O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}} + \frac{T^{1/4}}{\sqrt{s}}\right).$$

For Method 2, similar arguments give

$$\max_{i\le s,t\le T}\|(\widehat{\mathbf{b}}_i^{(2)})'\widehat{\mathbf{f}}_t^{(2)}-\mathbf{b}_i'\mathbf{f}_t\| = O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}}+\frac{T^{1/4}}{\sqrt{p}}\right).$$

In the oracle setting, where the factors are known, we have

$$\begin{aligned}\max_{i\le s,t\le T}\|(\widehat{\mathbf{b}}_i^{o})'\mathbf{f}_t-\mathbf{b}_i'\mathbf{f}_t\| = \max_{i\le s,t\le T}\|\widehat{\mathbf{b}}_i^{o}-\mathbf{b}_i\|\|\mathbf{f}_t\| &= O_P\left(\sqrt{\log T}\max_{i\le s}\|\widehat{\mathbf{b}}_i^{o}-\mathbf{b}_i\|\right)\\ &= O_P\left((\log T)^{1/2}\sqrt{\frac{\log s}{T}}\right).\end{aligned}$$

□

Lemma A.2. *Let* $\mathbf{H}_1=\widehat{\mathbf{V}}_1^{-1}\widehat{\mathbf{F}}^{(1)\prime}\mathbf{F}\mathbf{B}_S'\tilde{\mathbf{\Sigma}}_{u,S}^{-1}\mathbf{B}_S/T$ *and* $\mathbf{H}_2=\widehat{\mathbf{V}}_2^{-1}\widehat{\mathbf{F}}^{(2)\prime}\mathbf{F}\mathbf{B}'\tilde{\mathbf{\Sigma}}_u^{-1}\mathbf{B}/T$*, where* $\widehat{\mathbf{V}}_1$ *is the diagonal matrix of the largest* K *eigenvalues of* $\mathbf{Y}_S'\tilde{\mathbf{\Sigma}}_{u,S}^{-1}\mathbf{Y}_S/T$ *and* $\widehat{\mathbf{V}}_2$ *is the diagonal matrix of the largest* K *eigenvalues of* $\mathbf{Y}'\tilde{\mathbf{\Sigma}}_u^{-1}\mathbf{Y}/T$*. Under conditions of Lemma 1,* $\|\mathbf{H}_1\|=O_P(1)$ *and* $\|\mathbf{H}_2\|=O_P(1)$*.*

Proof of Lemma A.2. Since $\mathbf{\Sigma}_{u,S}$ is a submatrix of $\mathbf{\Sigma}_u$, it follows from condition (v) that $\lambda_{\min}(\mathbf{\Sigma}_{u,S}^{-1})\ge c_2^{-1}$. By Proposition 4.1 of Bai and Liao (2013), $\|\tilde{\mathbf{\Sigma}}_{u,S}^{-1}-\mathbf{\Sigma}_{u,S}^{-1}\|=o_P(1)$. Therefore, with probability tending to 1, $\|\tilde{\mathbf{\Sigma}}_{u,S}^{-1}\|\ge 1/(2c_2)$. Then,

$$T^{-1}\mathbf{Y}_S'\tilde{\mathbf{\Sigma}}_{u,S}^{-1}\mathbf{Y}_S = T^{-1}\mathbf{Y}_S'(\tilde{\mathbf{\Sigma}}_{u,S}^{-1}-(1/2c_2)\mathbf{I})\mathbf{Y}_S+1/(2c_2T)\mathbf{Y}_S'\mathbf{Y}_S.$$

Under the pervasive condition (i), it follows from Lemma C.4 of Fan et al. (2013) that the Kth largest eigenvalue of $T^{-1}\mathbf{Y}_S'\mathbf{Y}_S$ is larger than Ms. Since $T^{-1}\mathbf{Y}_S'(\tilde{\mathbf{\Sigma}}_{u,S}^{-1}-(1/2c_2)\mathbf{I})\mathbf{Y}_S$ is semi-positive definite, it follows from Weyl's inequality that

$$\lambda_K(T^{-1}\mathbf{Y}_S'\tilde{\mathbf{\Sigma}}_{u,S}^{-1}\mathbf{Y}_S)\ge\lambda_K(1/(2c_2T)\mathbf{Y}_S'\mathbf{Y}_S)\ge Ms/(2c_2).$$

Hence $\|\widehat{\mathbf{V}}_1^{-1}\|=O_P(s^{-1})$. Also, $\lambda_{\max}(\|\mathbf{F}'\mathbf{F}\|)=\lambda_{\max}(\|\sum_{t=1}^T\mathbf{f}_t\mathbf{f}_t'\|)=O_P(T)$. In addition, $\lambda_{\max}(\|\sum_{t=1}^T\widehat{\mathbf{f}}_t^{(1)}(\widehat{\mathbf{f}}_t^{(1)})'\|)=O_P(T)$, where the last equation follows from the constraint in (6). Then, $\|(\widehat{\mathbf{F}}^{(1)})'\mathbf{F}\|\le\|(\widehat{\mathbf{F}}^{(1)})'\widehat{\mathbf{F}}^{(1)}\|^{1/2}\|\mathbf{F}'\mathbf{F}\|^{1/2}=O_P(T)$. These results together with $\|\mathbf{B}_S'\tilde{\mathbf{\Sigma}}_{u,S}^{-1}\mathbf{B}_S\|=O(s)$ imply that $\|\mathbf{H}_1\|=O_P(1)$. Similarly, $\|\mathbf{H}_2\|=O_P(1)$. □

Lemma A.3. *(i)* $\|\mathbf{H}_1\mathbf{H}_1'-\mathbf{I}_K\|_F=O_P\left(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{s}}\right)$*; (ii)* $\|\mathbf{H}_2\mathbf{H}_2'-\mathbf{I}_K\|_F=O_P\left(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{p}}\right)$*. (iii)* $\|\mathbf{H}_1'\mathbf{H}_1-\mathbf{I}_K\|_F=O_P\left(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{s}}\right)$*; (iv)* $\|\mathbf{H}_2'\mathbf{H}_2-\mathbf{I}_K\|_F=O_P\left(\frac{1}{\sqrt{T}}+\frac{1}{\sqrt{p}}\right)$*.*

Proof of Lemma A.3. Let $\widehat{\mathrm{cov}}(\mathbf{H}_1\mathbf{f}_t)=\frac{1}{T}\sum_{t=1}^T(\mathbf{H}_1\mathbf{f}_t)(\mathbf{H}_1\mathbf{f}_t)'$. Then,

$$\|\mathbf{H}_1\mathbf{H}_1'-\mathbf{I}_K\|_F\le\underbrace{\|\mathbf{H}_1\mathbf{H}_1'-\widehat{\mathrm{cov}}(\mathbf{H}_1\mathbf{f}_t)\|_F}_{I_1}+\underbrace{\|\widehat{\mathrm{cov}}(\mathbf{H}_1\mathbf{f}_t)-\mathbf{I}_K\|_F}_{I_2}.$$

For I_1, we have $I_1 \leq \|\mathbf{H}_1\|^2 \|\mathbf{I}_K - \widehat{\mathrm{cov}}(\mathbf{f}_t)\|_F$, where $\widehat{\mathrm{cov}}(\mathbf{f}_t) = \frac{1}{T}\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t'$. It follows from Lemma C.3(i) of Fan et al. (2013) that $\|\mathbf{I}_K - \widehat{\mathrm{cov}}(\mathbf{f}_t)\|_F = O_P\left(1/\sqrt{T}\right)$. Then, $I_1 = O_P\left(1/\sqrt{T}\right)$, since $\|\mathbf{H}_1\| = O_P(1)$. For I_2, by the identifiability constraint in (6), $\frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{f}}_t^{(1)}\widehat{\mathbf{f}}_t^{(1)\prime} = \mathbf{I}_K$. Therefore,

$$
\begin{aligned}
I_2 &= \Big\| \frac{1}{T}\sum_{t=1}^T \mathbf{H}_1\mathbf{f}_t(\mathbf{H}_1\mathbf{f}_t)' - \frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{f}}_t^{(1)}\widehat{\mathbf{f}}_t^{(1)\prime} \Big\|_F \\
&\leq \Big\| \frac{1}{T}\sum_{t=1}^T (\mathbf{H}_1\mathbf{f}_t - \widehat{\mathbf{f}}_t^{(1)})(\mathbf{H}_1\mathbf{f}_t)' \Big\|_F + \Big\| \frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{f}}_t^{(1)}(\widehat{\mathbf{f}}_t^{(1)} - \mathbf{H}_1\mathbf{f}_t)' \Big\|_F \\
&\leq \left(\frac{1}{T}\sum_{t=1}^T \|\mathbf{H}_1\mathbf{f}_t - \widehat{\mathbf{f}}_t^{(1)}\|^2 \cdot \frac{1}{T}\sum_{t=1}^T \|\mathbf{H}_1\mathbf{f}_t\|^2 \right)^{1/2} + \left(\frac{1}{T}\sum_{t=1}^T \|\mathbf{H}_1\mathbf{f}_t - \widehat{\mathbf{f}}_t^{(1)}\|^2 \cdot \frac{1}{T}\sum_{t=1}^T \|\widehat{\mathbf{f}}_t^{(1)}\|^2 \right)^{1/2} \\
&= O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{s}} \right),
\end{aligned}
$$

where the last equality follows from Lemma A.5 and that $\|\mathbf{H}_1\mathbf{f}_t\| \leq \|\mathbf{H}_1\|\|\mathbf{f}_t\| = O_P(1)$ and $\|\widehat{\mathbf{f}}_t^{(1)}\| = O_P(1)$. Similarly, $\|\mathbf{H}_2\mathbf{H}_2' - \mathbf{I}_K\|_F = O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p}}\right)$.

(iii) Since $\|\mathbf{H}_1\mathbf{H}_1' - \mathbf{I}_K\|_F = O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{s}}\right)$ and $\|\mathbf{H}_1\| = O_P(1)$, we have $\|\mathbf{H}_1\mathbf{H}_1'\mathbf{H}_1 - \mathbf{H}_1\|_F = O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{s}}\right)$. Since $\mathbf{H}_1^{-1} = \mathbf{H}_1^{-1}(\mathbf{I}_K - \mathbf{H}_1\mathbf{H}_1' + \mathbf{H}_1\mathbf{H}_1')$, it follows Lemma A.3(i) that $\|\mathbf{H}_1^{-1}\| \leq \|\mathbf{H}_1^{-1}\| O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{s}}\right) + \|\mathbf{H}_1'\|$. Hence, $\|\mathbf{H}_1^{-1}\| = O_P(1)$. Left multiplying $\mathbf{H}_1\mathbf{H}_1'\mathbf{H}_1 - \mathbf{H}_1$ by $\mathbf{H}_1^{-1}$ gives $\|\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K\|_F = O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{s}}\right)$. Similarly, $\|\mathbf{H}_2'\mathbf{H}_2 - \mathbf{I}_K\|_F = O_P\left(\frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p}}\right)$. □

Lemma A.4. *Let* $\mathbf{C}_1 = \widehat{\mathbf{B}}_1 - \mathbf{B}_S\mathbf{H}_1'$ *and* $\mathbf{C}_2 = \widehat{\mathbf{B}}_2 - \mathbf{B}_S\mathbf{H}_2'$, *where* $\widehat{\mathbf{B}}_1$, $\widehat{\mathbf{B}}_2$, *and* $\mathbf{B}_S$ *are defined in Section 4.*
(i) $\|\mathbf{C}_1\|_F^2 = O_P(sw_1^2)$, $\|\mathbf{C}_2\|_F^2 = O_P(sw_2^2)$; $\|\mathbf{C}_1\mathbf{C}_1'\|_{\mathbf{\Sigma}_S}^2 = O_P(sw_1^4)$, $\|\mathbf{C}_2\mathbf{C}_2'\|_{\mathbf{\Sigma}_S}^2 = O_P(sw_2^4)$.
(ii) $\|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2 = O_P(m_s^2 w_1^2)$; $\|\widehat{\mathbf{\Sigma}}_{u,S}^{(2)} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2 = O_P(m_s^2 w_2^2)$.
(iii) $\|\mathbf{B}_S\mathbf{H}_1'\mathbf{C}_1'\|_{\mathbf{\Sigma}_S}^2 = O_P(w_1^2)$; $\|\mathbf{B}_S\mathbf{H}_2'\mathbf{C}_2'\|_{\mathbf{\Sigma}_S}^2 = O_P(w_2^2)$.
(iv) $\|\mathbf{B}_S(\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K)\mathbf{B}_S'\|_{\mathbf{\Sigma}_S}^2 = O_P\left(\frac{1}{sT} + \frac{1}{s^2}\right)$; $\|\mathbf{B}_S(\mathbf{H}_2'\mathbf{H}_2 - \mathbf{I}_K)\mathbf{B}_S'\|_{\mathbf{\Sigma}_S}^2 = O_P\left(\frac{1}{sT} + \frac{1}{sp}\right)$.

Proof of Lemma A.4. (i) We have $\|\mathbf{C}_1\|_F^2 \leq s(\max_{i\leq s}\|\widehat{\mathbf{b}}_i^{(1)} - \mathbf{H}\mathbf{b}_i\|)^2 = O_P(sw_1^2)$. By the general result that for any matrix $\mathbf{A}$, $\|\mathbf{A}\|_{\mathbf{\Sigma}_S}^2 = s^{-1}\|\mathbf{\Sigma}_S^{-1/2}\mathbf{A}\mathbf{\Sigma}_S^{-1/2}\|_F^2 = O_P(s^{-1}\|\mathbf{A}\|_F^2)$, we have $\|\mathbf{C}_1'\mathbf{C}_1\|_{\mathbf{\Sigma}_S}^2 = O_P(s^{-1}\|\mathbf{C}_1\|_F^4) = O_P(sw_1^4)$. Similarly, $\|\mathbf{C}_2\|_F^2 = O_P(sw_2^2)$ and $\|\mathbf{C}_2\mathbf{C}_2'\|_{\mathbf{\Sigma}_S}^2 = O_P(sw_2^4)$.

(ii) By Lemma 3,

$$
\|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)} - \mathbf{\Sigma}_{u,S}\|_{\mathbf{\Sigma}_S}^2 = O_P\left(s^{-1}\|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)} - \mathbf{\Sigma}_{u,S}\|_F^2\right) = O_P\left(\|\widehat{\mathbf{\Sigma}}_{u,S}^{(1)} - \mathbf{\Sigma}_{u,S}\|^2\right) = O_P\left(m_s^2 w_1^2\right).
$$

Similar results can be shown for $\|\widehat{\boldsymbol{\Sigma}}_{u,S}^{(2)} - \boldsymbol{\Sigma}_{u,S}\|_{\boldsymbol{\Sigma}_S}$.

(iii) By adapt the proof of Theorem 2 in Fan et al. (2008), we have that $\|\mathbf{B}_S'\boldsymbol{\Sigma}_S^{-1}\mathbf{B}_S\| = O(1)$. Hence,

$$\begin{aligned}\|\mathbf{B}_S\mathbf{H}_1'\mathbf{C}_1'\|_{\boldsymbol{\Sigma}_S}^2 &= s^{-1}\mathrm{tr}(\mathbf{H}_1'\mathbf{C}_1'\boldsymbol{\Sigma}_S^{-1}\mathbf{C}_1\mathbf{H}_1\mathbf{B}_S'\boldsymbol{\Sigma}_S^{-1}\mathbf{B}_S)\\ &\le s^{-1}\|\mathbf{H}_1\|^2\|\mathbf{B}_S'\boldsymbol{\Sigma}_S^{-1}\mathbf{B}_S\|\|\boldsymbol{\Sigma}_S^{-1}\|\|\mathbf{C}_1\|_F^2\\ &= O_P\left(s^{-1}\|\mathbf{C}_1\|_F^2\right) = O_P\left(w_1^2\right).\end{aligned}$$

Similarly, $\|\mathbf{B}_S\mathbf{H}_2'\mathbf{C}_2'\|_{\boldsymbol{\Sigma}_S} = O_P\left(w_2^2\right)$.

(iv) We have

$$\begin{aligned}\|\mathbf{B}_S(\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K)\mathbf{B}_S'\|_{\boldsymbol{\Sigma}_S}^2 &= s^{-1}\mathrm{tr}((\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K)\mathbf{B}_S'\boldsymbol{\Sigma}_S^{-1}\mathbf{B}_S(\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K)\mathbf{B}_S'\boldsymbol{\Sigma}_S^{-1}\mathbf{B}_S)\\ &\le s^{-1}\|\mathbf{H}_1'\mathbf{H}_1 - \mathbf{I}_K\|_F^2\|\mathbf{B}_S'\boldsymbol{\Sigma}_S^{-1}\mathbf{B}_S\|^2 = O_P\left(\frac{1}{sT} + \frac{1}{s^2}\right).\end{aligned}$$

Similarly, $\|\mathbf{B}_S(\mathbf{H}_2'\mathbf{H}_2 - \mathbf{I}_K)\mathbf{B}_S'\|_{\boldsymbol{\Sigma}_S}^2 = O_P\left(\frac{1}{sT} + \frac{1}{sp}\right)$. □

Lemma A.5. *Under conditions of Lemma 1,*

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t^{(1)} - \mathbf{H}_1\mathbf{f}_t\|^2 = O_P\left(1/s + 1/T\right),$$
$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t^{(2)} - \mathbf{H}_2\mathbf{f}_t\|^2 = O_P\left(1/p + 1/T\right).$$

Proof of Lemma A.5. Without loss of generality, we only prove the result for general p. Again, we write $\widehat{\mathbf{f}}_t^{(2)}$ as $\widehat{\mathbf{f}}_t$, $\mathbf{H}_2$ as $\mathbf{H}$ and $\widehat{\mathbf{V}}_2$ as $\widehat{\mathbf{V}}$ for notational simplicity. By (A.2),

$$\frac{1}{T}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t\|^2 \le c\|\widehat{\mathbf{V}}^{-1}\|^2\sum_{j=1}^{7}\frac{1}{T}\sum_{t=1}^{T}G_{jt}^2,$$

where c is a positive constant and G_{jt} is the jth summand on the right hand side of (A.2). By Lemma A.6 (iv) of Bai and Liao (2013), $\frac{1}{T}\sum_{i=1}^{T}G_{1t}^2 = o_P\left(1/p + 1/T\right)$. By Lemma A.10 (i) and (iii) of Bai and Liao (2013), $\frac{1}{T}\sum_{t=1}^{T}G_{2t}^2 = O_P\left(1/T\right)$ and $\frac{1}{T}\sum_{t=1}^{T}G_{3t}^2 = O_P\left(1/T\right)$. By Lemma A.6 (iii), (v) and (vi) of Bai and Liao (2013), $\frac{1}{T}\sum_{t=1}^{T}G_{4t}^2 = o_P(1/p)$, $\frac{1}{T}\sum_{t=1}^{T}G_{6t}^2 = o_P(1/p)$ and $\frac{1}{T}\sum_{t=1}^{T}G_{7t}^2 = o_P(1/p)$. Finally, by Lemma A.11 (ii) of Bai and Liao (2013), $\frac{1}{T}\sum_{t=1}^{T}G_{5t}^2 = O_P(1/p)$. Therefore, the dominating terms are G_{2t}, G_{3t} and G_{5t}, which together give the rate of $O_P\left(1/p + 1/T\right)$. □

Lemma A.6. *With probability tending to 1,*
(i) $\lambda_{\min}(\mathbf{I}_K + (\mathbf{B}_S\mathbf{H}_1')'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1') \ge cs$, $\lambda_{\min}(\mathbf{I}_K + (\mathbf{B}_S\mathbf{H}_2')'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_2') \ge cs$, $\lambda_{\min}(\mathbf{I}_K + \mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S) \ge cs$;
(ii) $\lambda_{\min}(\mathbf{I}_K + \widehat{\mathbf{B}}_1'(\widehat{\boldsymbol{\Sigma}}_{u,S}^{(1)})^{-1}\widehat{\mathbf{B}}_1) \ge cs$, $\lambda_{\min}(\mathbf{I}_K + \widehat{\mathbf{B}}_2'(\widehat{\boldsymbol{\Sigma}}_{u,S}^{(2)})^{-1}\widehat{\mathbf{B}}_2) \ge cs$, $\lambda_{\min}(\mathbf{I}_K + \widehat{\mathbf{B}}_o'(\widehat{\boldsymbol{\Sigma}}_{u,S}^{o})^{-1}\widehat{\mathbf{B}}_o) \ge cs$;
(iii) $\lambda_{\min}((\mathbf{H}_1'\mathbf{H}_1)^{-1} + \mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S) \ge cs$, $\lambda_{\min}((\mathbf{H}_2'\mathbf{H}_2)^{-1} + \mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S) \ge cs$.

Proof of Lemma A.6. By Lemma A.3, with probability tending to one, $\lambda_{\min}(\mathbf{H}_1\mathbf{H}_1')$ is bounded away from 0. Therefore,

$$\lambda_{\min}(\mathbf{I}_K + (\mathbf{B}_S\mathbf{H}_1')'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1') \geq \lambda_{\min}(\mathbf{H}_1\mathbf{B}_S'\boldsymbol{\Sigma}_{u,S}^{-1}\mathbf{B}_S\mathbf{H}_1')$$
$$\geq \lambda_{\min}(\boldsymbol{\Sigma}_{u,S}^{-1})\lambda_{\min}(\mathbf{B}_S'\mathbf{B}_S)\lambda_{\min}(\mathbf{H}_1\mathbf{H}_1') \geq cs.$$

Similar results hold for the other two statements. The results in (ii) follow from (i) and (A.6). The statement (iii) follows from a similar argument as $\mathbf{H}_1\mathbf{H}_1'$ and $\mathbf{H}_2\mathbf{H}_2'$ are positive semi-definite. □

Lemma A.7. *[Theorem A.1 of Fan et al. (2013)] Let $\widehat{u}_{it}$ be defined as in step ii. of Method 1 in Section 4. Under conditions (iv), (v), if there is a sequence $a_T = o(1)$ so that $\max_{i\leq p}\frac{1}{T}\sum_{t=1}^{T}|u_{it}-\widehat{u}_{it}|^2 = O_P(a_T^2)$ and $\max_{i\leq p,t\leq T}|u_{it}-\widehat{u}_{it}| = o_P(1)$, then the adaptive thresholding estimator $\widehat{\boldsymbol{\Sigma}}_u$ with $\omega(p) = \sqrt{(\log p)/T} + a_T$ satisfies that $\|\widehat{\boldsymbol{\Sigma}}_u - \boldsymbol{\Sigma}_u\| = O_P(m_p[\omega(p)]^{1-q})$. If further $m_p[\omega(p)]^{1-q} = o(1)$, then $\widehat{\boldsymbol{\Sigma}}_u$ is invertible with probability approaching one, and $\|\widehat{\boldsymbol{\Sigma}}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\| = O_P(m_p[\omega(p)]^{1-q})$.*

References

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

Bai, J. and Liao, Y. (2013). Statistical inferences using large estimated covariances for panel data and factor models. *arXiv:1307.2662*.

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

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

Shao, J. (2003). *Mathematical Statistics.* Springer-Verlag.

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