Online Supplement for "Inference for Approximate Factor Models: Random Missing and Cross Validation"

Sainan Jin[a], Ke Miao[b], and Liangjun Su[c]

[a] *School of Economics, Singapore Management University, Singapore*

[b] *School of Economics, Fudan University, Shanghai, China*

[c] *School of Economics and Management, Tsinghua University, Beijing, China*

This online supplement is composed of two parts. Section C contains the proofs of the technical lemmas and theorems in Appendices A and B. Section D provides some additional simulation results.

C Proofs of the Technical Lemmas and Theorems in Appendices A and B

Proof of Lemma A.1. From the principal component analysis (PCA), we have the identity $\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F}=\tilde{F}\tilde{D}$. Pre-multiplying both sides by $T^{-1}\tilde{F}'$ and using the normalization $T^{-1}\tilde{F}'\tilde{F}=I_R$ yield $T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F}=\tilde{D}$. The lemma follows provided $D=$plim$\tilde{D}$, which we show below.

Noting that $\tilde{X}=(F^0\Lambda^{0\prime}+\varepsilon)\circ G$, we have

$$\begin{aligned}
\tilde{D} &= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F}\\
&= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime}+\varepsilon)\circ G\right]\left[(F^0\Lambda^{0\prime}+\varepsilon)\circ G\right]'\tilde{F}\\
&= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ G\right]\left[(F^0\Lambda^{0\prime})\circ G\right]'\tilde{F}+T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}(\varepsilon\circ G)(\varepsilon\circ G)'\tilde{F}\\
&\quad +T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ G\right](\varepsilon\circ G)'\tilde{F}+T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}(\varepsilon\circ G)'\left[(F^0\Lambda^{0\prime})\circ G\right]'\tilde{F}\\
&\equiv D_{NT,1}+D_{NT,2}+D_{NT,3}+D_{NT,4},\ \text{say.}
\end{aligned}$$

We first study $D_{NT,1}$. Noting that $E(G)=q\mathbf{1}_{T\times N}$ with $\mathbf{1}_{T\times N}$ being a $T\times N$ matrix of ones, we make the following decomposition

$$\begin{aligned}
D_{NT,1} &= T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ G\right]\left[(F^0\Lambda^{0\prime})\circ G\right]'\tilde{F}\\
&= \frac{q^2}{\tilde{q}^2}\frac{\tilde{F}'F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\tilde{F}}{T}+T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\tilde{F}\\
&\quad +\frac{q}{\tilde{q}^2}T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right](F^0\Lambda^{0\prime})\tilde{F}+\frac{q}{\tilde{q}^2}T^{-1}\tilde{F}'\left(NT\tilde{q}^2\right)^{-1}(F^0\Lambda^{0\prime})\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\tilde{F}\\
&\equiv D_{NT,11}+D_{NT,12}+D_{NT,13}+D_{NT,14}
\end{aligned}$$

where $\tilde{G}=G-E(G)$. By the i.i.d. property of g_{it}, we can readily show that $\tilde{q}/q-1=O_P((NT)^{-1/2})$. By Lemma A.3(ii) in Bai (2003), $\frac{\tilde{F}'F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\tilde{F}}{T}\overset{p}{\to}D$. This result can be strengthened to $||\frac{\tilde{F}'F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\tilde{F}}{T}-D||=O_P(\delta_{NT}^{-1})$ under our assumptions. Then $\|D_{NT,11}-D\|=O_P(\delta_{NT}^{-1})$.

For $D_{NT,12}$, we have

$$\begin{aligned}
\|D_{NT,12}\|_{\text{sp}} &= \left(NT\tilde{q}^2\right)^{-1}\lambda_{\max}\left(T^{-1}\tilde{F}'\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\tilde{F}\right) \\
&\leq \operatorname{tr}\left(T^{-1}\tilde{F}'\tilde{F}\right)\left(NT\tilde{q}^2\right)^{-1}\lambda_{\max}\left(\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]'\right) \\
&= R\left(NT\tilde{q}^2\right)^{-1}\left\|(F^0\Lambda^{0\prime})\circ\tilde{G}\right\|_{\text{sp}}^2
\end{aligned}$$

where the last equality follows from the fact that $\operatorname{tr}(T^{-1}\tilde{F}'\tilde{F})$ $=\operatorname{tr}(I_R)=R$. Let $c_{\lambda,F}=\max_{i,t}\left|\lambda_i^{0\prime}F_t^0\right|$ and $Z=[(F^0\Lambda^{0\prime})\circ\tilde{G}]/c_{\lambda,F}$. Let Z_{it} denote a typical element of $Z:Z_{it}=\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)/c_{\lambda,F}$. By construction, $\max_{i,t}|Z_{it}|\leq 1$. We want to apply Lemma C.4 by conditioning on $\mathcal{F}=\sigma\left\{F^0,\Lambda^0\right\}$, the sigma-field generated by F^0 and Λ^0. By straightforward moment calculations

$$\begin{aligned}
c_1 &\equiv \max_i\sqrt{\sum_{t=1}^T E\left(Z_{it}^2|\mathcal{F}\right)}=\max_i\sqrt{\sum_{t=1}^T\frac{\left(\lambda_i^{0\prime}F_t^0\right)^2}{c_{\lambda,F}^2}E\left(g_{it}-q\right)^2} \\
&= \frac{\sqrt{q\left(1-q\right)}}{c_{\lambda,F}}\max_i\sqrt{\lambda_i^{0\prime}F^{0\prime}F^0\lambda_i^0}\leq\frac{c_{\lambda,N}\left\|F^{0\prime}F^0\right\|^{1/2}}{c_{\lambda,F}},
\end{aligned}$$

and

$$\begin{aligned}
c_2 &\equiv \max_t\sqrt{\sum_{i=1}^N E\left(Z_{it}^2|\mathcal{F}\right)}=\max_t\sqrt{\sum_{i=1}^N\frac{\left(\lambda_i^{0\prime}F_t^0\right)^2}{c_{\lambda,F}^2}E\left(g_{it}-q\right)^2} \\
&= \frac{\sqrt{q\left(1-q\right)}}{c_{\lambda,F}}\max_t\sqrt{F_t^{0\prime}\Lambda^{0\prime}\Lambda^0F_t^0}\leq\frac{c_{F,T}\left\|\Lambda^{0\prime}\Lambda^0\right\|^{1/2}}{c_{\lambda,F}},
\end{aligned}$$

where $c_{\lambda,N}=\max_i\left\|\lambda_i^0\right\|$ and $c_{F,T}=\max_t\left\|F_t^0\right\|$. It follows that

$$\left\|(F^0\Lambda^{0\prime})\circ\tilde{G}\right\|_{\text{sp}}=O_P\left(\max\left\{c_{\lambda,N}\left\|F^{0\prime}F^0\right\|^{1/2},c_{F,T}\left\|\Lambda^{0\prime}\Lambda^0\right\|^{1/2},c_{\lambda,F}\log(N\vee T)\right\}\right).$$

This result, in conjunction with the fact $\left\|F^{0\prime}F^0\right\|=O_P\left(T\right),\left\|\Lambda^{0\prime}\Lambda^0\right\|=O_P\left(N\right),c_{\lambda,N}=O_P\left(N^{\gamma_1/4}\right),$ $c_{F,T}=O_P\left(T^{\gamma_2/4}\right)$, and $c_{\lambda,F}=O_P\left((NT)^{1/4}\right)$ under our moment conditions on λ_i^0 and F_t^0 in Assumption A.1, implies that

$$\begin{aligned}
\|D_{NT,12}\| &\leq \sqrt{R}\|D_{NT,12}\|_{\text{sp}}=O_P\left(\frac{1}{NT}\max\left\{c_{\lambda,N}^2T,c_{F,T}^2N,c_{\lambda,F}^2\left[\log(N\vee T)\right]^2\right\}\right) \\
&\leq O_P\left\{\frac{1}{NT}\max\left\{N^{\gamma_1/2}T,T^{\gamma_2/2}N,(NT)^{1/2}\left[\log(N\vee T)\right]^2\right\}\right\}=O_P(\delta_{NT}^{-(2-\gamma)})
\end{aligned}$$

where $\gamma=\gamma_1\vee\gamma_2$. Then $\|D_{NT,13}\|=\|D_{NT,14}\|\leq\{\|D_{NT,11}\|\,\|D_{NT,12}\|\}^{1/2}=O_P(\delta_{NT}^{-(1-\gamma/2)})$ by the matrix version of Cauchy-Schwarz (CS) inequality. Therefore we have $\|D_{NT,1}-D\|=O_P(\delta_{NT}^{-(1-\gamma/2)})$.

Noting that $D_{NT,2}$ is positive semidefinite (p.s.d.), we have

$$\begin{aligned}
\|D_{NT,2}\|_{\text{sp}} &\leq \left(NT\tilde{q}^2\right)^{-1}\operatorname{tr}\left(T^{-1}\tilde{F}'\left(\varepsilon\circ G\right)\left(\varepsilon\circ G\right)'\tilde{F}\right)\leq\operatorname{tr}\left(T^{-1}\tilde{F}'\tilde{F}\right)\left(NT\tilde{q}^2\right)^{-1}\lambda_{\max}\left(\left(\varepsilon\circ G\right)\left(\varepsilon\circ G\right)'\right) \\
&= R\left(NT\tilde{q}^2\right)^{-1}\left\|\varepsilon\circ G\right\|_{\text{sp}}^2,
\end{aligned}$$

where the first inequality follows from the fact that $\|A\|_{\mathrm{sp}} = \lambda_{\max}(A) \le \mathrm{tr}(A)$ for any p.s.d. symmetric matrix A, the second inequality follows because $\mathrm{tr}(A'BA) \le \mathrm{tr}(A'A)\,\lambda_{\max}(B)$ for any symmetric p.s.d. matrix B and conformable matrix A, the equality follows because $\mathrm{tr}(T^{-1}\tilde{F}'\tilde{F}) = \mathrm{tr}(I_R) = R$. Note that

$$\|\varepsilon \circ G\|_{\mathrm{sp}} \le \left\|\varepsilon \circ \tilde{G}\right\|_{\mathrm{sp}} + \|\varepsilon \circ E(G)\|_{\mathrm{sp}} = \left\|\varepsilon \circ \tilde{G}\right\|_{\mathrm{sp}} + q\,\|\varepsilon\|_{\mathrm{sp}}.$$

By Assumption A.2(i), $\|\varepsilon\|_{\mathrm{sp}} = O_P(\sqrt{N}+\sqrt{T})$. As in the analysis of $\left\|(F^0\Lambda^{0\prime}) \circ \tilde{G}\right\|_{\mathrm{sp}}$, we can readily apply Lemma B.1 by conditioning on ε to obtain with high probability

$$\left\|\varepsilon \circ \tilde{G}\right\|_{\mathrm{sp}} = O_P\left(\max\left\{\sqrt{N}, \sqrt{T}, \max_{i,t}|\varepsilon_{it}|\log(N \vee T)\right\}\right) \le O_P\left(\sqrt{N}+\sqrt{T}+(NT)^{1/4}\log(N\vee T)\right).$$

It follows that $\|D_{NT,2}\| \le \sqrt{R}\,\|D_{NT,2}\|_{\mathrm{sp}} \le (NT)^{-1} O_P\left(N+T+(NT)^{1/2}[\log(N\vee T)]^2\right) = o_P(\delta_{NT}^{-(2-\gamma)})$ and $\|D_{NT,2}\| \le \sqrt{R}\,\|D_{NT,2}\|_{\mathrm{sp}} = o_P(\delta_{NT}^{-(2-\gamma)})$ and $\|D_{NT,3}\| = \|D_{NT,4}\| \le \{\|D_{NT,1}\|\,\|D_{NT,2}\|\}^{1/2} = o_P(\delta_{NT}^{-(1-\gamma/2)})$ by the CS inequality.

In sum, we have $\left\|\tilde{D} - D\right\| = O_P(\delta_{NT}^{-(1-\gamma/2)})$. ■

Proof of Lemma A.2. (i) From the method of PCA, we have

$$\left(NT\tilde{q}^2\right)^{-1}\tilde{X}\tilde{X}'\tilde{F} = \tilde{F}\tilde{D}. \tag{C.1}$$

Using $\tilde{X} = \left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G$ and $G = E(G)+\tilde{G} = q\mathbf{1}_{T\times N}+\tilde{G}$, we make the following decomposition

$$\begin{aligned}
&\tilde{X}\tilde{X}' \\
=\;& \left[\left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G\right]\left[\left(F^0\Lambda^{0\prime}+\varepsilon\right)\circ G\right]' \\
=\;& \left[(F^0\Lambda^{0\prime})\circ G\right]\left[(F^0\Lambda^{0\prime})\circ G\right]' + (\varepsilon\circ G)(\varepsilon\circ G)' + \left[(F^0\Lambda^{0\prime})\circ G\right](\varepsilon\circ G)' + (\varepsilon\circ G)'\left[(F^0\Lambda^{0\prime})\circ G\right]' \\
=\;& q^2F^0\Lambda^{0\prime}\Lambda^0F^{0\prime} + d_{NT},
\end{aligned} \tag{C.2}$$

where $d_{NT} = \left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]' + q(F^0\Lambda^{0\prime})\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]' + q\left[(F^0\Lambda^{0\prime})\circ\tilde{G}\right]\Lambda^0F^{0\prime} + (\varepsilon\circ G)(\varepsilon\circ G)' + \left[(F^0\Lambda^{0\prime})\circ G\right](\varepsilon\circ G)' + (\varepsilon\circ G)'\left[(F^0\Lambda^{0\prime})\circ G\right]'$. Premultiplying both sides of (C.1) by $\left(\frac{1}{N}\Lambda^{0\prime}\Lambda^0\right)^{1/2}\times\frac{1}{T}F^{0\prime}$ and plugging (C.2) yield

$$\frac{q^2}{\tilde{q}^2}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\tilde{F}}{T}\right) + \bar{d}_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\tilde{F}}{T}\right)\tilde{D}, \tag{C.3}$$

where $\bar{d}_{NT} = \frac{1}{\tilde{q}^2}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\frac{1}{T}F^{0\prime}d_{NT}\tilde{F}$. Following the analysis of D_{NT}'s in the proof of Lemma A.1, we can readily show that $\left\|\bar{d}_{NT}\right\| = O_P(\delta_{NT}^{-(1-\gamma/2)})$. Letting $B_{NT} = \frac{q^2}{\tilde{q}^2}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}$ and $R_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\tilde{F}}{T}\right)$, we can write (C.3) as follows: $[B_{NT}+\bar{d}_{NT}R_{NT}^{-1}]R_{NT} = R_{NT}\tilde{D}$. Hence, each column of R_{NT} is non-standardized eigenvector of the matrix $B_{NT}+\bar{d}_{NT}R_{NT}^{-1}$. Let $\breve{D}_{NT}$ be a diagonal matrix consisting of the diagonal elements of $R_{NT}'R_{NT}$. Denote the standardized eigenvector $\Upsilon_{NT} = R_{NT}\breve{D}_{NT}^{-1/2}$. Hence, we have $[B_{NT} + \bar{d}_{NT}R_{NT}^{-1}]\Upsilon_{NT} = \Upsilon_{NT}\tilde{D}^{-1}$. That is, $\tilde{D}$ contains the eigenvalues of $B_{NT}+\bar{d}_{NT}R_{NT}^{-1}$ with the corresponding normalized eigenvectors contained in Υ_{NT}. It is trivial to show that with high probability

$$\left\|B_{NT} + \bar{d}_{NT}R_{NT}^{-1} - B\right\| = O_P(\delta_{NT}^{-(1-\gamma/2)}), \tag{C.4}$$

where B denotes the probability of B_{NT}, i.e., $B = \Sigma_{\Lambda^0}^{1/2}\Sigma_{F^0}\Sigma_{\Lambda^0}^{1/2}$.

Let Υ denote the probability limit of Υ_{NT}. Note that $\Upsilon' = \Upsilon^{-1}$ by normalization. By (C.4) and the eigenvector perturbation theory that requires distinctness of eigenvalues (see, e.g., Steward and Sun (1990, Ch. V), and Allez and Bouchaud (2013)), $\|\Upsilon_{NT} - \Upsilon\| = O_P(\delta_{NT}^{-(1-\gamma/2)})$ by (C.4) and Assumption A.1(iv). This, in conjunction with the definition of R_{NT}, implies that $\frac{F^{0\prime}\tilde{F}}{T} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1/2} R_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1/2} \Upsilon_{NT}\breve{D}_{NT}^{1/2}$ satisfies $\left\|\frac{F^{0\prime}\tilde{F}}{T} - \Sigma_{\Lambda^0}^{-1/2}\Upsilon D^{1/2}\right\| = O_P(\delta_{NT}^{-(1-\gamma/2)})$. The result follows by noticing that $Q' = \Sigma_{\Lambda^0}^{-1/2}\Upsilon D^{1/2}$.

(ii) By Lemma A.1, (i) and Assumption A.1(ii) , we have

$$\begin{aligned}\tilde{H} &= \left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\left(T^{-1}F^{0\prime}\tilde{F}\right)\tilde{D}^{-1} = \Sigma_{\Lambda^0}(\Sigma_{\Lambda^0}^{-1/2}\Upsilon D^{1/2})D^{-1} + O_P(\delta_{NT}^{-(1-\gamma/2)}) \\ &= \Sigma_{\Lambda^0}^{1/2}\Upsilon D^{-1/2} + O_P(\delta_{NT}^{-(1-\gamma/2)}) = Q^{-1} + O_P(\delta_{NT}^{-(1-\gamma/2)}).\end{aligned}$$

(iii) The proof follows closely that of Lemma B.1 in Bai (2003) and Theorem 2.1 and thus omitted. The major difference is that we now use the decomposition in (A.1) and the fact that g_{it} are i.i.d. Bernoulli(q) and independent of F^0, Λ^0 and ε.

(iv) The proof is analogous to that of Theorem 2.1 and thus omitted.

(v) The claim follows from (iv) provided that we can show that $\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t - \tilde{H}'F_t^0)F_t^{0\prime}g_{it} = O_P\left(\delta_{NT}^{-2}\right)$. The proof of the latter result follows closely that of Theorem 2.1 (or Lemma B.2 in Bai (2003)) and thus omitted.

(vi) By (v), the claim follows provided that $\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t - \tilde{H}'F_t^0)F_t^{0\prime} = O_P\left(\delta_{NT}^{-2}\right)$. We can prove the latter result by using analogous arguments as used in the proof of Theorem 2.1 and Lemma B.2 in Bai (2003).

(vii) Using $\tilde{F}_t = (\tilde{F}_t - \tilde{H}'F_t^0) + \tilde{H}'F_t^0$, we make the following decomposition

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'(g_{it}-q) &= \tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H}(g_{it}-q) + \frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t - \tilde{H}'F_t^0)(\tilde{F}_t - \tilde{H}'F_t^0)'(g_{it}-q) \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t - \tilde{H}'F_t^0)F_t^{0\prime}\tilde{H}'(g_{it}-q) + \tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0(\tilde{F}_t - \tilde{H}'F_t^0)'(g_{it}-q) \\ &\equiv \sum_{l=1}^{4}d_{lt}.\end{aligned}$$

By Theorem 2.1 and Lemma A.2(iv), $d_{2t} = O_P\left(\delta_{NT}^{-2}\right)$. By Lemma A.2(vi)-(vii), $d_{3t} = O_P\left(\delta_{NT}^{-2}\right)$ and $d_{4t} = O_P\left(\delta_{NT}^{-2}\right)$. Then $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'(g_{it}-q) = \tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H}(g_{it}-q) + O_P\left(\delta_{NT}^{-2}\right)$.

(viii) As in (vii), we can also show that $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t' = \tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H} + O_P\left(\delta_{NT}^{-2}\right)$. This, in conjunction with the fact that $\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t' = I_R$, implies that $\tilde{H}'\frac{1}{T}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\tilde{H} = I_R + O_P\left(\delta_{NT}^{-2}\right)$. Premultiplying and postmultiplying both sides by $(\tilde{H}')^{-1}$ and $\tilde{H}^{-1}$ in order yields $\frac{1}{T}F^{0\prime}F^0 = (\tilde{H}\tilde{H}')^{-1} + O_P\left(\delta_{NT}^{-2}\right)$. It follows that $\tilde{H}\tilde{H}' = (\frac{1}{T}F^{0\prime}F^0)^{-1} + O_P\left(\delta_{NT}^{-2}\right)$. ■

A Cautionary Note. We can prove Lemmas A.3-A.5 for $\ell = 1$ based on the results in Theorems 2.1-2.2. When these lemmas hold for $\ell = 1$, Theorems 2.3-2.4 also hold for $\ell = 1$. With the results in Lemmas A.3-A.5 and Theorems 2.3-2.4 for $\ell = 1$, we can prove them to hold for $\ell = 2$. This procedure is repeated until convergence. Since the verification of Lemma A.3 for $\ell = 1$ is different from the general case with $\ell \geq 2$, we first prove it for $\ell = 1$ in detail and then prove it for $\ell \geq 2$ after we prove Lemmas A.4-A.5.

Proof of Lemma A.3 ($\ell = 1$)**.** (i) Noting that $\hat{\phi}_{F,t}^{(0)} = \hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0\frac{1}{Nq}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0\left(g_{it} - q\right)\right]$, $\max_t\left\|\hat{\phi}_{F,t}^{(0)}\right\| \leq O_P\left(1\right)\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0\left(g_{it} - q\right)\right]\right\| = O_P\left((N/\ln N)^{-1/2}\right)$ by Lemmas A.1-A.2 and Assumption A.5(i). Similarly, $\max_i\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\| \leq O_P\left(1\right)\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0\left[\varepsilon_{it}g_{it} + \lambda_i^{0\prime}F_t^0\left(g_{it} - q\right)\right]\right\| = O_P\left((T/\ln T)^{-1/2}\right)$ by Lemmas A.1-A.2 and Assumption A.5(ii).

(ii) By the decomposition in (A.1),

$$\hat{r}_{F,t}^{(0)} = \hat{F}_t^{(0)} - \hat{H}^{(0)\prime}F_t^0 - \hat{\phi}_{F,t}^{(0)} = a_{1t} + a_{2t} + a_{4t} + a_{5t} + a_{7t} + (a_{3t} + a_{6t} - \hat{\phi}_{F,t}^{(0)}).$$

Following the proof of Theorem 2.2(i) and using Assumption A.5 and the fact that $\max_t\left\|F_t^0\right\| = O_P\left(T^{\gamma_1/4}\right)$, it is easy to show that

$$\begin{aligned}
\max_t\|a_{1t}\| &= O_P\left(T^{-1/2}\delta_{NT}^{-1} + T^{-1+\gamma_1/4}\right), \ \max_t\|a_{2t}\| = O_P\left(\delta_{NT}^{-2}\ln N\right), \\
\max_t\|a_{lt}\| &= O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\right) \text{ for } l = 4, 5, \\
\max_t\|a_{7t}\| &= O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}\right),
\end{aligned}$$

and $\max_t\left\|a_{3t} + a_{6t} - \hat{\phi}_{F,t}^{(0)}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$. It follows that $\max_t\left\|\hat{r}_{F,t}^{(0)}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})$. For $\hat{r}_{\Lambda,i}^{(0)}$, we have

$$\hat{r}_{\Lambda,i}^{(0)} = \hat{\lambda}_i^{(0)} - (\hat{H}^{(0)})^{-1}\lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(0)} = B_{2i} + B_{3i} + B_{5i} + (B_{1i} + B_{4i} - \hat{\phi}_{\Lambda,i}^{(0)}),$$

where B_{2i}'s are defined in the proof of Theorem 2.2(ii). Following the proof of Theorem 2.2(ii) and using the fact that $\max_i\left\|\lambda_i^0\right\| = O_P\left(N^{\gamma_2/4}\right)$, $\max_i\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^2 = O_P\left(1\right)$, and $\tilde{q} - q = O_P((NT)^{-1/2})$ we have by Theorem 2.1 and Lemma A.2

$$\max_i\|B_{2i}\| = O_P\left(\delta_{NT}^{-2}\ln N\right), \ \max_i\|B_{3i}\| = O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right), \ \max_i\|B_{5i}\| = O_P\left((NT)^{-1/2}N^{\gamma_2/4}\right),$$

and $\max_i\left\|B_{1i} + B_{4i} - \hat{\phi}_{\Lambda,i}^{(0)}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$. It follows that $\max_i\left\|\hat{r}_{\Lambda,i}^{(0)}\right\| = O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right)$.

(iii) By (i) and the fact that $\max_t \left\|F_t^0\right\| = O_P\left(T^{\gamma_1/4}\right)$ and $\max_i \left\|\lambda_i^0\right\| = O_P\left(N^{\gamma_2/4}\right)$, we have

$$\begin{aligned}
& \max_{i,t} \left\|\eta_{1,it}^{(0)}\right\| \\
= & \max_{i,t} \left\|F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)} + \lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)} + F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}\right\| \\
\leq & \left\|\hat{H}^{(0)}\right\| \max_t \|F_t^0\| \left\{\max_i \left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\| + \hat{r}_{\Lambda,i}^{(\ell)}\right\} + \left\|(\hat{H}^{(0)\prime})^{-1}\right\| \max_i \|\lambda_i^0\| \left\{\max_t \left\|\hat{\phi}_{F,t}^{(0)}\right\| + \max_t \left\|\hat{r}_{F,t}^{(\ell)}\right\|\right\} \\
= & O_P(T^{\gamma_1/4}((T/\ln T)^{-1/2} + N^{\gamma_2/4}\delta_{NT}^{-2}\ln N)) + O_P(T^{\gamma_1/4}((T/\ln T)^{-1/2} + T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}) \\
= & O_P(\delta_{NT}^{-1+\gamma/2}\ln N).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\max_{i,t}\left\|\eta_{2,it}^{(0)}\right\| &= \max_{i,t}\left\|\hat{\phi}_{\Lambda,i}^{(0)\prime}\hat{\phi}_{F,t}^{(0)} + \hat{\phi}_{\Lambda,i}^{(0)\prime}\hat{r}_{F,t}^{(0)} + \hat{\phi}_{F,t}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)} + \hat{r}_{\Lambda,i}^{(0)\prime}\hat{r}_{F,t}^{(0)}\right\| \\
&\leq O_P\left((N/\ln N)^{-1/2}(T/\ln T)^{-1/2}\right) + O_P(T/\ln T)^{-1/2}\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}\right)) \\
&\quad + O_P\left((N/\ln N)^{-1/2}N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right) + O_P\left(\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}\right)\right) \\
&= O_P(\delta_{NT}^{-2}\ln N).
\end{aligned}$$

(iv) Note that

$$\begin{aligned}
& [\hat{H}^{(0)\prime}]^{-1}\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it} \\
= & \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\left[\varepsilon_{is}g_{is} + \lambda_i^{0\prime}F_s^0(g_{is}-q)\right]\varepsilon_{it}g_{it} \\
= & \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0 E(\varepsilon_{is}\varepsilon_{it})g_{is}g_{it} + \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\left[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})\right]g_{is}g_{it} \\
& + \frac{q}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0F_s^{0\prime}\lambda_i^0(g_{is}-q)\varepsilon_{it}g_{it} + \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0F_s^{0\prime}\lambda_i^0(g_{is}-q)\varepsilon_{it}(g_{it}-q) \\
= & O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N + \delta_{NT}^{-2}\ln N + \delta_{NT}^{-2}\ln N\right) = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right).
\end{aligned}$$

Then $\max_t \left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it}\right\| = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$.

Observe that $\left\|\hat{H}^{(0)}\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\| = \left\|\hat{H}^{(0)}\hat{H}^{(0)\prime}\right\|\left\|\frac{1}{NTq}\sum_{i=1}^N\sum_{s=1}^T \lambda_i^0F_s^{0\prime}\left[\varepsilon_{is}g_{is} + F_s^{0\prime}\lambda_i^0(g_{is}-q)\right]\bar{g}_{it}\right\|$. Using $\bar{g}_{it} = (1-q) - (g_{it}-q)$, we have

$$\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T \lambda_i^0F_s^{0\prime}\left[\varepsilon_{is}g_{is} + F_s^{0\prime}\lambda_i^0(g_{is}-q)\right]\bar{g}_{it} \\
= & \frac{1-q}{NT}\sum_{i=1}^N\sum_{s=1}^T \lambda_i^0F_s^{0\prime}\varepsilon_{is}g_{is} + \frac{1-q}{NT}\sum_{i=1}^N\sum_{s=1}^T \lambda_i^0F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q) \\
& - \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T \lambda_i^0F_s^{0\prime}\varepsilon_{is}g_{is}(g_{it}-q) - \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T \lambda_i^0F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q)(g_{it}-q).
\end{aligned}$$

It is easy to show that the first two terms are $O_P(\delta_{NT}^{-2})$ by Chebyshev inequality. The third term is $O_P(\delta_{NT}^{-2}\ln N)$ by Assumption A.3(iii). For the fourth term, we have

$$\begin{aligned}
& \max_t \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\lambda_i^0 F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q)(g_{it}-q)\right\| \\
= & \max_t \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T}\lambda_i^0 F_s^{0\prime}F_s^{0\prime}\lambda_i^0(g_{is}-q)(g_{it}-q)\right\| + \frac{1}{NT}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2 \max_t \left\|F_t^0\right\|^2 \\
= & O_P(\delta_{NT}^{-2}\ln N) + O_P(T^{-1+\gamma_1/2}).
\end{aligned}$$

Then $\left\|\hat{H}^{(0)}\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\| = O_P\left(T^{-1+\gamma_1/2}+\delta_{NT}^{-2}\ln N\right).$

Noting that $\hat{r}_{\Lambda,i}^{(0)} = \hat{\lambda}_i^{(0)} - (\hat{H}^{(0)})^{-1}\lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(0)} = B_{2i}+B_{3i}+B_{5i}+(B_{1i}+B_{4i}-\hat{\phi}_{\Lambda,i}^{(0)})$, we have

$$\max_t \left\|\frac{1}{N}\sum_{i=1}^{N}\hat{r}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\| \leq \max_t \left\|\frac{1}{N}\sum_{i=1}^{N}\left[B_{2i}+B_{3i}+B_{5i}+(B_{1i}+B_{4i}-\hat{\phi}_{\Lambda,i}^{(0)})\right]\lambda_i^{0\prime}\bar{g}_{it}\right\|,$$

where B_{2i}'s are defined in the proof of Theorem 2.2(ii). Using $\bar{g}_{it} = (1-q)+(g_{it}-q)$,

$$\begin{aligned}
\max_t \left\|\frac{1}{N}\sum_{i=1}^{N}B_{2i}\lambda_i^{0\prime}\bar{g}_{it}\right\| &= \frac{\left\|\hat{H}^{(0)}\right\|}{\tilde{q}}\max_t \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}g_{is}\lambda_i^{0\prime}\bar{g}_{it}\right\| \\
&\leq O_P(1)\left\{\left\|\frac{1-q}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}g_{is}\lambda_i^{0\prime}\right\| + \max_t \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\varepsilon_{is}g_{is}\lambda_i^{0\prime}(g_{it}-q)\right\|\right\} \\
&= O_P((NT)^{-1/2}+(NT)^{-1/2}\ln N).
\end{aligned}$$

In addition,

$$\begin{aligned}
\max_t \left\|\frac{1}{N}\sum_{i=1}^{N}B_{3i}\lambda_i^{0\prime}\bar{g}_{it}\right\| &= \max_t \frac{1}{NT\tilde{q}}\left\|\sum_{i=1}^{N}\sum_{s=1}^{T}\hat{F}_s^{(0)}\left(\hat{H}^{(0)\prime}F_s^0-\hat{F}_s^{(0)}\right)'(\hat{H}^{(0)\prime})^{-1}\lambda_i^0 g_{is}\lambda_i^{0\prime}\bar{g}_{it}\right\| \\
&\leq O_P(1)\max_i \left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(0)}\left(\hat{H}^{(0)\prime}F_s^0-\hat{F}_s^{(0)}\right)' g_{is}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right),
\end{aligned}$$

and $\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}B_{5i}\lambda_i^{0\prime}\bar{g}_{it}\right\| \leq \frac{1}{\tilde{q}}\left|q-\tilde{q}\right|\left\|[\hat{H}^{(0)\prime}]^{-1}\right\|\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2 = O_P((NT)^{-1/2})$. Lastly, noting that the difference lies between $B_{1i}+B_{4i}$ and $\hat{\phi}_{\Lambda,i}^{(0)}$ is controlled by $\left|\tilde{q}-q\right|$, we can readily show that $\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}(B_{1i}+B_{4i}-\hat{\phi}_{\Lambda,i}^{(0)})\lambda_i^{0\prime}\bar{g}_{it}\right\| = O_P((NT)^{-1/2})$. In sum, we have $\max_t\left\|\hat{H}^{(0)\prime}\frac{1}{N}\sum_{i=1}^{N}\hat{r}_{\Lambda,i}^{(0)}\lambda_i^{0\prime}\bar{g}_{it}\right\|$ $= O_P\left(\delta_{NT}^{-2}\ln N\right)$.

(v) Noting that $\hat{\phi}_{F,t}^{(0)}=\hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0\frac{1}{Nq}\sum_{i=1}^N\lambda_i^0\left[\varepsilon_{it}g_{it}+\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]$, we have

$$\begin{aligned}\max_i\left\|\frac{1}{T}\sum_{t=1}^T\hat{\phi}_{F,t}^{(0)}F_t^{0\prime}\bar{g}_{it}\right\| &\leq O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0\left[\varepsilon_{jt}g_{jt}+\lambda_j^{0\prime}F_t^0\left(g_{jt}-q\right)\right]F_t^{0\prime}\bar{g}_{it}\right\| \\ &\leq O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\varepsilon_{jt}g_{jt}\bar{g}_{it}\right\| \\ &\quad+O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\lambda_j^{0\prime}F_t^0\left(g_{jt}-q\right)\bar{g}_{it}\right\| \\ &= O_P\left(\delta_{NT}^{-2}\ln N\right)+O_P(N^{-1+\gamma_2/2}).\end{aligned}$$

Analogously, by the decomposition in (A.1) we have $\frac{1}{T}\sum_{t=1}^T\hat{r}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}=\frac{1}{T}\sum_{t=1}^T[a_{1t}+a_{2t}+a_{4t}+a_{5t}+a_{7t}+(a_{3t}+a_{6t}-\hat{\phi}_{F,t}^{(0)})]F_t^{0\prime}\bar{g}_{it}$. Following the proof of Theorem 2.2(i) and using Assumption A.5 and the fact that $\max_i\left\|\lambda_i^0\right\|=O_P\left(N^{\gamma_2/4}\right)$, it is easy to show that $\max_i||\frac{1}{T}\sum_{t=1}^T a_{1t}F_t^{0\prime}\bar{g}_{it}||=O_P\left(T^{-1/2}\delta_{NT}^{-1}+T^{-1}\right)$, $\max_i||\frac{1}{T}\sum_{t=1}^T a_{2t}F_t^{0\prime}\bar{g}_{it}||\leq\max_t\left\|a_{2t}\right\|O_P(1)=O_P\left(\delta_{NT}^{-2}\ln N\right)$, $\max_i||\frac{1}{T}\sum_{t=1}^T a_{lt}F_t^{0\prime}\bar{g}_{it}||=O_P\left(\delta_{NT}^{-2}\ln N\right)$ for $l=4,5$, $\max_i||\frac{1}{T}\sum_{t=1}^T a_{7t}F_t^{0\prime}\bar{g}_{it}||=O_P\left(\delta_{NT}^{-2}\ln T+T^{-1}\right)$, and $\max_i||\frac{1}{T}\sum_{t=1}^T[a_{3t}+a_{6t}-\hat{\phi}_{F,t}^{(0)}]F_t^{0\prime}\bar{g}_{it}||=O_P\left(\delta_{NT}^{-2}\ln N\right)$. It follows that $\max_i||\frac{1}{T}\sum_{t=1}^T\hat{r}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}||=O_P\left(\delta_{NT}^{-2}\ln N\right)$.

(vi) Note that $\frac{1}{N}\sum_{i=1}^N\left\|\eta_{it}^{(0)}\right\|^2\leq\frac{2}{N}\sum_{i=1}^N\left\|\eta_{1,it}^{(0)}\right\|^2+\frac{2}{N}\sum_{i=1}^N\left\|\eta_{2,it}^{(0)}\right\|^2$, where the second term is bounded above by $O_P(\delta_{NT}^{-4}(\ln N)^2)$ by (iii). For the first term, we have

$$\begin{aligned}\max_t\frac{1}{N}\sum_{i=1}^N\left\|\eta_{1,it}^{(0)}\right\|^2 &\leq \max_t\frac{1}{N}\sum_{i=1}^N\left\|F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}+\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)}+\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)}+F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}\right\|^2 \\ &\leq 4\left\|\hat{H}^{(0)}\right\|\max_t\left\|F_t^0\right\|^2\frac{1}{N}\sum_{i=1}^N(\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2+\left\|\hat{r}_{\Lambda,i}^{(0)}\right\|^2) \\ &\quad+4\left\|[\hat{H}^{(0)\prime}]^{-1}\right\|\left\{\max_t\left\|\hat{\phi}_{F,t}^{(0)}\right\|^2+\max_t\left\|\hat{r}_{F,t}^{(0)}\right\|^2\right\}\frac{1}{N}\sum_{i=1}^N\left\|\lambda_i^0\right\|^2 \\ &= O_P(T^{-1+\gamma_1/2}+N^{-1}\ln N).\end{aligned}$$

It follows that $\frac{1}{N}\sum_{i=1}^N\left\|\eta_{it}^{(0)}\right\|^2=O_P(T^{-1+\gamma_1/2}+N^{-1}\ln N)$. Similarly, we can show that $\max_t\frac{1}{T}\sum_{t=1}^T||\eta_{it}^{(0)}||^2=O_P(N^{-1+\gamma_2/2}+T^{-1}\ln N)$.

(vii) Let $\kappa_t=1+\left\|F_t^0\right\|^2$. It suffices to show that $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\kappa_t(\eta_{l,it}^{(0)})^2=O_P\left(\delta_{NT}^{-2}\right)$ for $l=1,2$. By (iii), $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\kappa_t(\eta_{2,it}^{(0)})^2\leq\max_{i,t}||\eta_{2,it}^{(0)}||^2\frac{1}{T}\sum_{t=1}^T\kappa_t=O_P(\delta_{NT}^{-4}(\ln N)^2)$. In addition, $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N\kappa_t(\eta_{1,it}^{(0)})^2\leq\frac{4}{NT}\sum_{t=1}^T\sum_{i=1}^N\kappa_t\{||F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}||^2+||\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)}||^2+||\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)}||^2+||F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}||^2\}\equiv4\left\{J_{1,1}+J_{1,2}+J_{1,3}+J_{1,4}\right\}$. For $J_{1,1}$, we have

$$J_{1,1}\leq\left\|\hat{H}^{(0)}\right\|^2\frac{1}{T}\sum_{t=1}^T\kappa_t\left\|F_t^0\right\|^2\frac{1}{N}\sum_{i=1}^N\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2=O_P\left(T^{-1}\right),$$

as we can readily show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^{2}=O_P\left(T^{-1}\right)$. For $J_{1,2}$, noting that $(\hat{H}^{(0)\prime})^{-1}\hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^0$ $=(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}$ and $\frac{1}{N}\Lambda^{0\prime}\Lambda^0-\Sigma_{\Lambda^0}=O\left(N^{-1/2}\right)$, we have

$$\begin{aligned} J_{1,2} &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t\left\|\lambda_i^{0\prime}(\frac{1}{N}\Lambda^{0\prime}\Lambda^0)^{-1}\frac{1}{Nq}\sum_{j=1}^{N}\lambda_j^0\left[\varepsilon_{jt}g_{ij}+\lambda_j^{0\prime}F_t^0\left(g_{jt}-q\right)\right]\right\|^2 \\ &\leq O_P\left(1\right)\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\left[\varepsilon_{it}g_{it}+\lambda_i^{0\prime}F_t^0\left(g_{it}-q\right)\right]\right\|^2=O_P\left(\delta_{NT}^{-2}\right). \end{aligned}$$

Similarly, we can show that $J_{1,l}=O_P\left(\delta_{NT}^{-2}\right)$ for $l=3,4$. Then $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{1,it}^{(0)})^2=O_P\left(\delta_{NT}^{-2}\right)$.

(viii) Note that $\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\eta_{is}^{(0)}\bar{g}_{is}=\sum_{l=1}^{2}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\eta_{l,is}^{(0)}\bar{g}_{is}\equiv\sum_{l=1}^{2}J_{2,l}$. For $J_{2,2}$, we can use the uniform bound in (iii) and show that $J_{2,l}=O_P\left(\delta_{NT}^{-2}\ln N\right)$. For $J_{2,1}$, we have $J_{2,1}=\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i^{0\prime}\left(F_t^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}+\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)}+\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,t}^{(0)}+F_t^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}\right)\bar{g}_{it}\equiv$ $\sum_{a=1}^{4}J_{2,1a}$. Let λ_{il}^0 and F_{sl}^0 denote the lth element of λ_i^0 and F_s^0, respectively. Let $J_{2,1a}\left(l,r\right)$ denote the (l,r)th element of $J_{2,1a}$ for $a=1,2$. Noting that $\bar{g}_{is}=(1-q)+(q-g_{is})$, we have

$$\begin{aligned} \left\|J_{2,11}\left(l,r\right)\right\| &= \left\|\frac{1}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(0)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\bar{g}_{it}\lambda_{il}^0\right\| \\ &\leq \left\|\frac{1}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(0)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\left(1-q\right)\lambda_{il}^0\right\|+\left\|\frac{1}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(0)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}(g_{it}-q)\lambda_{il}^0\right\| \\ &\equiv J_{2,1a}\left(l,r,1\right)+J_{2,1a}\left(l,r,2\right). \end{aligned}$$

First, $J_{2,1a}\left(l,r,1\right)\leq(1-q)\left\|\hat{H}^{(0)}\right\|\frac{1}{T}\sum_{t=1}^{T}\left\|F_t^0\right\|^2\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\lambda_{il}^0\right\|=O_P\left(1\right)\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\lambda_{il}^0\right\|=$ $O_P(\delta_{NT}^{-2})$ by the fact that $\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\lambda_{il}^0\right\|\leq\left\|\hat{H}^{(0)}\right\|\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{Tq}\sum_{t=1}^{T}F_t^0\left[\varepsilon_{it}g_{it}+F_t^{0\prime}\lambda_i^0(g_{it}-q)\right]\lambda_{il}^0\right\|$ $=O_P(\delta_{NT}^{-2})$ by Chebyshev inequality. For $J_{2,1a}\left(l,r,2\right)$, we have

$$\begin{aligned} J_{2,1a}\left(l,r,2\right) &= \left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{il}^0\hat{\phi}_{\Lambda,i}^{(0)\prime}\hat{H}^{(0)\prime}\left[\frac{1}{T}\sum_{t=1}^{T}F_t^0F_{tr}^0(g_{it}-q)\right]\right\| \\ &\leq \left\|\hat{H}^{(0)}\right\|\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0F_{tr}^0(g_{it}-q)\right\|^2\right\}^{1/2} \\ &= O_P\left(\delta_{NT}^{-1}\right)O_P(T^{-1/2}), \end{aligned}$$

as we can show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^2\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|^2=O_P(\delta_{NT}^{-2})$ and $\frac{1}{N}\sum_{i=1}^{N}E\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0F_{tr}^0(g_{it}-q)\right\|^2=$ $O(T^{-1})$. Then $J_{2,11}=O_P(\delta_{NT}^{-2})$. Similarly,

$$\begin{aligned} \left\|J_{2,12}\left(l,r\right)\right\| &= \left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_{ir}^0\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)}F_{tl}^0\bar{g}_{it}\right\| \\ &\leq \left\|\frac{1-q}{N}\sum_{i=1}^{N}\lambda_{ir}^0\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\frac{1}{T}\sum_{t=1}^{T}\hat{\phi}_{F,t}^{(0)}F_{tl}^0\right\|+\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_{ir}^0\lambda_i^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,t}^{(0)}F_{tl}^0\left(g_{it}-q\right)\right\| \\ &\equiv J_{2,12}\left(l,r,1\right)+J_{2,12}\left(l,r,2\right). \end{aligned}$$

Noting that $\frac{1}{T}\sum_{t=1}^{T}\hat{\phi}_{F,t}^{(0)}F_{tl}^{0}=\hat{D}^{(0)-1}\frac{1}{T}\hat{F}^{(0)\prime}F^{0}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{Nq}\sum_{i=1}^{N}\lambda_{i}^{0}\left[\varepsilon_{it}g_{it}+\lambda_{i}^{0\prime}F_{t}^{0}\left(g_{it}-q\right)\right]F_{tl}^{0}=O_{P}\left(1\right)$ $\times\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_{i}^{0}\left[\varepsilon_{it}g_{it}+\lambda_{i}^{0\prime}F_{t}^{0}\left(g_{it}-q\right)\right]F_{tl}^{0}=O_{P}\left(\delta_{NT}^{-2}\right)$, $\left\|J_{2,12}\left(l,r,1\right)\right\|=O_{P}\left(\delta_{NT}^{-2}\right)$. For $J_{2,12}\left(l,r,2\right)$, we have

$$\begin{aligned}
J_{2,12}\left(l,r,2\right) &= \left\|\frac{1}{T}\sum_{s=1}^{T}F_{sl}^{0}\hat{\phi}_{F,s}^{(0)\prime}(\hat{H}^{(0)})^{-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_{i}^{0}\lambda_{ir}^{0}\left(g_{is}-q\right)\right\| \\
&\leq \left\|(\hat{H}^{(0)})^{-1}\right\|\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|F_{s}^{0}\right\|^{2}\left\|\hat{\phi}_{F,s}^{(0)}\right\|^{2}\right\}^{1/2}\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{i}^{0}\lambda_{ir}^{0}\left(g_{is}-q\right)\right\|^{2}\right\}^{1/2} \\
&= O_{P}\left(\delta_{NT}^{-1}\right)O_{P}(N^{-1/2}).
\end{aligned}$$

So $J_{2,2}=O_{P}\left(\delta_{NT}^{-2}\right)$. Similarly, we can show that $J_{2,l}=O_{P}\left(\delta_{NT}^{-2}\right)$ for $l=3,4$. Then $\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_{s}^{0}\lambda_{i}^{0\prime}$ $\times\eta_{is}\bar{g}_{is}=O_{P}\left(\delta_{NT}^{-2}\right)$.

(ix) By (vi) and the fact that $\frac{1}{N}\sum_{i=1}^{N}E\left\|\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}\varepsilon_{is}g_{is}\right\|^{2}=O(T^{-1})$,

$$\begin{aligned}
\max_{t}\left\|\frac{1}{NT}\sum_{s=1}^{T}F_{s}^{0}\sum_{i=1}^{N}\eta_{it}^{(0)}\bar{g}_{it}\varepsilon_{is}g_{is}\right\| &= \max_{t}\left\|\frac{1}{N}\sum_{i=1}^{N}\eta_{it}^{(0)}\bar{g}_{it}\left(\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}\varepsilon_{is}g_{is}\right)\right\| \\
&\leq \left\{\max_{t}\frac{1}{N}\sum_{i=1}^{N}(\eta_{it}^{(0)})^{2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}\varepsilon_{is}g_{is}\right\|^{2}\right\}^{1/2} \\
&= O_{P}\left(T^{-1/2+\gamma_{1}/4}+(N/\ln N)^{-1/2}\right)O_{P}(T^{-1/2}) \\
&= O_{P}\left(T^{-1+\gamma_{1}/4}+(NT/\ln N)^{-1/2}\right).
\end{aligned}$$

(x) Note that $\frac{1}{NT}\sum_{s=1}^{T}F_{s}^{0}\sum_{i=1}^{N}\varepsilon_{it}g_{it}\eta_{is}^{(0)}\bar{g}_{is}=\sum_{l=1}^{2}\frac{1}{NT}\sum_{s=1}^{T}F_{s}^{0}\sum_{i=1}^{N}\varepsilon_{it}g_{it}\eta_{l,is}^{(0)}\bar{g}_{is}\equiv\sum_{l=1}^{2}J_{3,lt}$.We can readily bound $J_{3,2t}$ by $O_{P}(\delta_{NT}^{-2}\ln N)$ by using the uniform bound for $\eta_{2,is}^{(0)}$ in (iii). Next, $J_{3,1t}=\frac{1}{NT}\sum_{s=1}^{T}F_{s}^{0}\sum_{i=1}^{N}\varepsilon_{it}g_{it}[F_{s}^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}+\lambda_{i}^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{\phi}_{F,s}^{(0)}+\lambda_{i}^{0\prime}(\hat{H}^{(0)\prime})^{-1}\hat{r}_{F,s}^{(0)}+F_{t}^{0\prime}\hat{H}^{(0)\prime}\hat{r}_{\Lambda,i}^{(0)}]\bar{g}_{is}\equiv$ $\sum_{l=1}^{4}J_{3,1t}\left(l\right)$. Using $\bar{g}_{is}=(1-q)+(q-g_{is})$, the fact that $F_{s}^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}$ is a scalar and $\max_{t}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^{2}=$ $O_{P}\left(1\right)$, and (iv), we have

$$\begin{aligned}
&\max_{t}J_{3,1t}\left(1\right) \\
&= \max_{t}\left\|\frac{1}{NT}\sum_{s=1}^{T}F_{s}^{0}\sum_{i=1}^{N}\varepsilon_{it}g_{it}F_{s}^{0\prime}\hat{H}^{(0)}\hat{\phi}_{\Lambda,i}^{(0)}\bar{g}_{is}\right\| \\
&\leq \max_{t}\left\|\frac{1-q}{T}\sum_{s=1}^{T}F_{s}^{0}F_{s}^{0\prime}\hat{H}^{(0)}\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it}\right\|+\max_{t}\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}g_{it}\hat{\phi}_{\Lambda,i}^{(0)\prime}\hat{H}^{(0)\prime}\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}F_{s}^{0\prime}\left(g_{is}-q\right)\right\| \\
&\leq O_{P}\left(1\right)\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(0)}\varepsilon_{it}g_{it}\right\|+\max_{i}\left\|\hat{\phi}_{\Lambda,i}^{(0)}\right\|\left\{\max_{t}\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^{2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}F_{s}^{0\prime}\left(g_{is}-q\right)\right\|^{2}\right\}^{1/2} \\
&= O_{P}(T^{-1+\gamma_{1}/4}+N^{-1}\ln N)+O_{P}((T/\ln T)^{-1/2})O_{P}\left(1\right)O_{P}(T^{-1/2})=O_{P}(T^{-1+\gamma_{1}/4}+\delta_{NT}^{-2}\ln N).
\end{aligned}$$

For $J_{3,1t}(2)$, we have by (i) and the fact that $\max_s \left\| \frac{1}{N}\sum_{i=1}^N \lambda_i^0 \varepsilon_{it} g_{it} \bar{g}_{is} \right\| = O_P(N/\ln N)^{-1/2})$,

$$\begin{aligned}
\max_t J_{3,1t}(2) &= \max_t \left\| \frac{1}{T}\sum_{s=1}^T F_s^0 \hat{\phi}_{F,s}^{(0)\prime} [\hat{H}^{(0)}]^{-1} \frac{1}{N}\sum_{i=1}^N \lambda_i^0 \varepsilon_{it} g_{it} \bar{g}_{is} \right\| \\
&\leq \left\| [\hat{H}^{(0)}]^{-1} \right\| \max_t \left\| \hat{\phi}_{F,t}^{(0)} \right\| \max_s \left\| \frac{1}{N}\sum_{i=1}^N \lambda_i^0 \varepsilon_{it} g_{it} \bar{g}_{is} \right\| \frac{1}{T}\sum_{s=1}^T \left\| F_s^0 \right\| \\
&= O_P((N/\ln N)^{-1/2}) O_P(N/\ln N)^{-1/2}) = O_P\left(\delta_{NT}^{-2}\ln N\right).
\end{aligned}$$

Similarly, we can show that $J_{3,1t}(l) = O_P(\delta_{NT}^{-2}\ln N)$ for $l = 3, 4$. Then $\max_t \left\| \frac{1}{NT}\sum_{s=1}^T F_s^0 \sum_{i=1}^N \varepsilon_{it} g_{it} \eta_{is}^{(0)} \bar{g}_{is} \right\|$ $= O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$. ■

Proof of Lemma A.4 ($\ell = 1$ and $\ell > 1$)**.** (i) From the PCA, we have the identity $(NT)^{-1}\hat{X}^{(\ell)}\hat{X}^{(\ell)\prime}\hat{F}^{(\ell)}$ $= \hat{F}^{(\ell)}\hat{D}^{(\ell)}$. Pre-multiplying both sides by $T^{-1}\hat{F}^{(\ell)\prime}$ and using the normalization $T^{-1}\hat{F}^{(\ell)\prime}\hat{F}^{(\ell)} = I_R$ yield $T^{-1}\hat{F}^{(\ell)\prime}(NT)^{-1}\hat{X}\hat{X}'\hat{F}^{(\ell)} = \hat{D}^{(\ell)}$. Let $\varepsilon^{(\ell)}$ be the $T \times N$ matrix with (t,i)th element given by $\varepsilon_{it}^{(\ell-1)} = \varepsilon_{it} g_{it} + \eta_{it}^{(\ell-1)} \bar{g}_{it}$. Noting that $\hat{X}^{(\ell)} = F^0\Lambda^{0\prime} + \varepsilon_{it}^{(\ell)}$, we have

$$\begin{aligned}
\hat{D}^{(\ell)} &= T^{-1}\hat{F}^{(\ell)\prime}(NT)^{-1}(F^0\Lambda^{0\prime} + \varepsilon^{(\ell)})(F^0\Lambda^{0\prime} + \varepsilon^{(\ell)})'\hat{F}^{(\ell)} \\
&= T^{-1}\hat{F}^{(\ell)\prime}(NT)^{-1}\left\{F^0\Lambda^{0\prime}\Lambda^0 F^{0\prime} + \varepsilon^{(\ell)}\varepsilon^{(\ell)\prime} + F^0\Lambda^{0\prime}\varepsilon^{(\ell)\prime} + \varepsilon^{(\ell)}\Lambda^0 F^{0\prime}\right\}\hat{F}^{(\ell)} \\
&\equiv \hat{D}_1^{(\ell)} + \hat{D}_2^{(\ell)} + \hat{D}_3^{(\ell)} + \hat{D}_4^{(\ell)}.
\end{aligned}$$

The result follows if we show that (1) $\hat{D}_1^{(\ell)} = D + O_P(\delta_{NT}^{-1}\ln N)$ and (2) $\hat{D}_l^{(\ell)} = O_P(\delta_{NT}^{-1}\ln N)$ for $l = 2, 3, 4$. Following the proof of Lemma A.1(i) in Su and Wang (2017), $\hat{D}_1^{(\ell)} = \frac{\hat{F}^{(\ell)\prime}F^0}{N}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^0\hat{F}^{(\ell)}}{T} = D + O_P\left(\delta_{NT}^{-1}\right)$. Noting that $\varepsilon^{(\ell)} = \varepsilon \circ G + \eta^{(\ell-1)} \circ \bar{G}$ where $\bar{G} = \mathbf{1}_{T\times N} - G$ and $\eta^{(\ell-1)}$ has (t,i)th element given by $\eta_{it}^{(\ell-1)}$,

$$\begin{aligned}
\left\| \hat{D}_2^{(\ell)} \right\|_{\text{sp}} &= (NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}(\varepsilon \circ G + \eta^{(\ell-1)} \circ \bar{G})(\varepsilon \circ G + \eta^{(\ell-1)} \circ \bar{G})'\hat{F}^{(\ell)}\right) \\
&\leq 2(NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}(\varepsilon \circ G)(\varepsilon \circ G)'\hat{F}^{(\ell)}\right) \\
&\quad + 2(NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}(\eta^{(\ell-1)} \circ \bar{G})(\eta^{(\ell-1)} \circ \bar{G})'\hat{F}^{(\ell)}\right).
\end{aligned}$$

Following the analysis of $D_{NT,2}$ in the proof of Lemma A.1, we can show that the first term is $O_P(\delta_{NT}^{-2}[\log(N \vee T)]^2)$. For the second term, it suffices to use Lemma A.3(vii) to obtain the following rough probability bound

$$2(NT)^{-1}\operatorname{tr}\left(T^{-1}\hat{F}^{(\ell)\prime}\eta^{(\ell-1)}\eta^{(\ell-1)\prime}\hat{F}^{(\ell)}\right) \leq 2T^{-1}\left\|\hat{F}^{(\ell)}\right\|^2 (NT)^{-1}\left\|\eta^{(\ell-1)}\right\|^2 = O_P(\delta_{NT}^{-2}).$$

It follows that $\left\|\hat{D}_2\right\| \leq R^{1/2}\left\|\hat{D}_2\right\|_{\text{sp}} = O_P(\delta_{NT}^{-2}(\ln N)^2)$. By the CS inequality, $\left\|\hat{D}_3^{(\ell)}\right\| = \left\|\hat{D}_4^{(\ell)}\right\| \leq \left\{\left\|\hat{D}_1^{(\ell)}\right\|\left\|\hat{D}_2^{(\ell)}\right\|\right\}^{1/2} = O_P(\delta_{NT}^{-1}\ln N))$. In sum, we have $\hat{D}^{(\ell)} = D + O_P(\delta_{NT}^{-1}\ln N)$.

(ii) The proof is analogous to that of Lemma A.2(i) with obvious modifications.

(iii) The proof is analogous to that of Lemma A.2(ii) with obvious modifications.

(iv) The proof follows from that of Lemma B.3 in Bai (2003).

(v) Note that $\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^{(\ell)}-\hat{H}^{(\ell)\prime}F_t^0)\varepsilon_{it}^{(\ell)}=\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^{(\ell)}-\hat{H}^{(\ell)\prime}F_t^0)\varepsilon_{it}g_{it}+\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^{(\ell)}-\hat{H}^{(\ell)\prime}F_t^0)\eta_{it}^{(\ell)}\bar{g}_{it}$. Following the proof of Lemma A.7(v) in Su and Wang (2017), we can show that the first term is $O_P\left(\delta_{NT}^{-2}\ln N\right)$ uniformly in t. By Theorem 2.3 and Lemma A.3(vi), the Frobenius norm of the second term is bounded above by above by$\{\frac{1}{T}||\hat{F}^{(\ell)}-F^0\hat{H}^{(\ell)}||^2\}^{1/2}\{\max_i\frac{1}{T}\sum_{t=1}^{T}(\eta_{it}^{(\ell)})^2\}^{1/2}=\delta_{NT}^{-1}O_P(N^{-1/2+\gamma_2/4}+(T/\ln N)^{-1/2})$. It follows that $\frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^{(\ell)}-\hat{H}^{(\ell)\prime}F_t^0)\varepsilon_{it}^{(\ell)}=O_P(N^{-1/2+\gamma_2/4}\delta_{NT}^{-1}+\delta_{NT}^{-2}\ln N)$. ■

Proof of Lemma A.5 ($\ell=1$ and $\ell>1$). (i) Note that $\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}^{(\ell)}=\beta_{F,t}+\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\eta_{it}^{(\ell-1)}\bar{g}_{it}$ where

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\eta_{it}^{(\ell-1)}\bar{g}_{it}=\sum_{l=1}^{2}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\eta_{l,it}^{(\ell-1)}\bar{g}_{it}\equiv\sum_{l=1}^{2}K_{1,lt}.$$

By Lemma A.3(iii), $\max_t\|K_{1,2t}\|\leq\max_{i,t}\left\|\eta_{2,it}^{(\ell-1)}\right\|\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|=O_P(\delta_{NT}^{-2}\ln N)$. For $K_{1,1t}$, we make the following decomposition

$$\begin{aligned}K_{1,1t}&=\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\left[F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)}+\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}+F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\right]\bar{g}_{it}\\&\equiv K_{1,1t}(1)+K_{1,1t}(2)+K_{1,1t}(3)+K_{1,1t}(4).\end{aligned}$$

For $K_{1,1t}(1)$, we apply Lemma A.3(iv) to obtain $\max_t\|K_{1,1t}(1)\|=O_P\left(T^{\gamma_1/4}\delta_{NT}^{-2}\ln N\right)$. For $K_{1,1t}(2)$, we have uniformly in t

$$\begin{aligned}K_{1,1t}(2)&=\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)}\bar{g}_{it}\\&=(1-q)\frac{1}{N}\Lambda^{0\prime}\Lambda^0(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)}+\left(\frac{1}{N}\sum_{i=1}^{N}(g_{it}-q)\lambda_i^0\lambda_i^{0\prime}\right)(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)}\\&=(1-q)\left[[\hat{D}^{(\ell-1)}]^{-1}\frac{1}{T}\hat{F}^{(\ell-1)\prime}F^0\right]^{-1}\hat{\phi}_{F,t}^{(\ell-1)}+O_P(N^{-1}\ln N),\end{aligned}$$

where the second equality follows from the use of $\bar{g}_{it}=(1-q)-(g_{it}-q)$, the third equality holds by (i), the fact that $\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}(g_{it}-q)\right\|=O_P\left(N^{-1/2}\ln N\right)$, and the definition of $\hat{H}^{(\ell-1)}$. In addition, we have by (ii) and (iv)

$$\begin{aligned}\max_t\|K_{1,1t}(3)\|&=\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}\bar{g}_{it}\right\|\\&\leq\max_t\left\|\hat{r}_{F,t}^{(\ell-1)}\right\|O_P\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\lambda_i^{0\prime}\right\|\right)=O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T+T^{-1+3\gamma_1/4})\end{aligned}$$

and

$$\begin{aligned}\max_t \|K_{1,1t}(4)\| &= \max_t \left\| F_t^{0\prime} \hat{H}^{(\ell-1)\prime} \frac{1}{N} \sum_{i=1}^{N} \hat{r}_{\Lambda,i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| \\ &\leq \max_t \|F_t^{0\prime}\| \max_t \left\| \hat{H}^{(\ell-1)\prime} \frac{1}{N} \sum_{i=1}^{N} \hat{r}_{\Lambda,i}^{(\ell-1)} \lambda_i^{0\prime} \bar{g}_{it} \right\| = O_P(T^{\gamma_1/4} \delta_{NT}^{-2} \ln N).\end{aligned}$$

It follows that uniformly in t, $\frac{1}{N}\sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it}^{(\ell)} = \beta_{F,t} + (1-q)\left[[\hat{D}^{(\ell-1)}]^{-1} \frac{1}{T} \hat{F}^{(\ell-1)\prime} F^0\right]^{-1} \hat{\phi}_{F,t}^{(\ell-1)} + O_P(T^{\gamma_1/4}\delta_{NT}^{-2} \ln T + T^{-1+3\gamma_1/4})$ and

$$\hat{\phi}_{F,t}^{(\ell)} = [\hat{D}^{(\ell)}]^{-1} \frac{1}{T} \hat{F}^{(\ell)\prime} F^0 \frac{1}{N} \sum_{i=1}^{N} \lambda_i^0 \varepsilon_{it}^{(\ell)} = D^{-1} Q \beta_{F,t} + (1-q) \hat{\phi}_{F,t}^{(\ell-1)} + O_P(T^{\gamma_1/4}\delta_{NT}^{-2} \ln T + T^{-1+3\gamma_2/4}).$$

(ii) Note that $\frac{1}{T}\sum_{t=1}^{T} F_t^0 \varepsilon_{it}^{(\ell)} = \beta_{\Lambda,i} + \frac{1}{T}\sum_{t=1}^{T} F_t^0 \eta_{it}^{(\ell-1)} \bar{g}_{it}$ where $\frac{1}{T}\sum_{t=1}^{T} F_t^0 \eta_{it}^{(\ell-1)} \bar{g}_{it} = \sum_{l=1}^{2} \frac{1}{T}\sum_{t=1}^{T} F_t^0 \eta_{l,it}^{(\ell-1)} \bar{g}_{it} \equiv \sum_{l=1}^{2} K_{2i,l}$. By Lemma A.3(iii), we can show that $\max_i \|K_{2i,2}\| = O_P\left(\delta_{NT}^{-2} \ln N\right)$. Using the decomposition $\bar{g}_{it} = (1-q) + (g_{it} - q)$ and Lemma A.2, we can readily show that

$$\begin{aligned}K_{2i,1} &= \frac{1}{T}\sum_{t=1}^{T} F_t^0 \left[F_t^{0\prime} \hat{H}^{(\ell-1)} \hat{\phi}_{\Lambda,i}^{(\ell-1)} + \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{\phi}_{F,t}^{(\ell-1)} + \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{r}_{F,t}^{(\ell-1)} + F_t^{0\prime} \hat{H}^{(\ell-1)\prime} \hat{r}_{\Lambda,i}^{(\ell-1)}\right] \bar{g}_{it} \\ &\equiv K_{2i,1}(1) + K_{2i,1}(2) + K_{2i,1}(3) + K_{2i,1}(4).\end{aligned}$$

For $K_{2i,1}(1)$, we have that uniformly in i,

$$\begin{aligned}K_{2i,1}(1) &= \frac{1-q}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \hat{H}^{(\ell-1)} \hat{\phi}_{\Lambda,i}^{(\ell-1)} + \frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \hat{H}^{(\ell-1)} \hat{\phi}_{\Lambda,i}^{(\ell-1)} (g_{it} - q) \\ &= (1-q) \frac{1}{T} F^{0\prime} F^0 \hat{H}^{(\ell-1)} \hat{\phi}_{\Lambda,i}^{(\ell-1)} + O_P((NT)^{-1/2} \ln N) \\ &= (1-q) [\hat{H}^{(\ell-1)\prime}]^{-1} \hat{\phi}_{\Lambda,i}^{(\ell-1)} + O_P((NT)^{-1/2} \ln N),\end{aligned}$$

where the second equality follows from the fact that

$$\begin{aligned}\left\| \frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \hat{H}^{(\ell-1)} \hat{\phi}_{\Lambda,i}^{(\ell-1)} (g_{it} - q) \right\| &\leq O_P\left(\max_i \left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|\right) \max_i \left\| \frac{1}{T}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} (g_{it} - q) \right\| \\ &= O_P((N/\ln N)^{-1/2}) O_P((T/\ln T)^{-1/2})\end{aligned}$$

and the last equality follows because $\frac{1}{T} F^{0\prime} F^0 = [\hat{H}^{(\ell-1)} \hat{H}^{(\ell-1)\prime}]^{-1} + O_P(\delta_{NT}^{-2})$. By Lemma A.3(v) and (ii)

$$\begin{aligned}\max_i \|K_{2i,1}(2)\| &= \max_i \left\| \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \frac{1}{T}\sum_{t=1}^{T} \hat{\phi}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| \\ &\leq O_P\left(\max_i \|\lambda_i^0\|\right) \max_i \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{\phi}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| = N^{\gamma_2/4} O_P\left(\delta_{NT}^{-2} \ln N + N^{-1+\gamma_2/2}\right),\end{aligned}$$

$$\begin{aligned}
\max_i \|K_{2i,1}(3)\| &= \max_i \left\| \frac{1}{T}\sum_{t=1}^T F_t^0 \lambda_i^{0\prime} (\hat{H}^{(\ell-1)\prime})^{-1} \hat{r}_{F,t}^{(\ell-1)} \bar{g}_{it} \right\| \\
&\leq O_P\left(\max_i \|\lambda_i^0\|\right) \left\| \frac{1}{T}\sum_{t=1}^T \hat{r}_{F,t}^{(\ell-1)} F_t^{0\prime} \bar{g}_{it} \right\| = N^{\gamma_2/4} O_P\left(\delta_{NT}^{-2} \ln N\right), \text{ and} \\
\max_i \|K_{2i,1}(4)\| &= \max_i \left\| \frac{1}{T}\sum_{t=1}^T F_t^0 F_t^{0\prime} \hat{H}^{(\ell-1)\prime} \hat{r}_{\Lambda,i}^{(\ell-1)} \bar{g}_{it} \right\| \\
&\leq O_P\left(\max_i \left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\|\right) \frac{1}{T}\sum_{t=1}^T \left\|F_t^0\right\|^2 = O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N).
\end{aligned}$$

It follows that uniformly in i, $\frac{1}{T}\sum_{t=1}^T F_t^0 \varepsilon_{it}^{(\ell)} = \beta_{\Lambda,i} + (1-q)[\hat{H}^{(\ell-1)\prime}]^{-1}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + O_P\left(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N\right)$ and

$$\hat{\phi}_{\Lambda,i}^{(\ell)} = \hat{H}^{(\ell-1)\prime}\frac{1}{T}\sum_{t=1}^T F_t^0 \varepsilon_{it}^{(\ell)} = (Q')^{-1}\beta_{\Lambda,i} + (1-q)\hat{\phi}_{\Lambda,i}^{(\ell-1)} + O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N + N^{-1+3\gamma_2/4}). \blacksquare$$

Proof of Lemma A.3 ($\ell \geq 2$). The proof relies on the fact that Lemmas A.3-A.5 and Theorems 2.3-2.4 hold for $\ell - 1$.

(i) By Lemma A.5(i)-(ii),

$$\begin{aligned}
\max_t \left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\| &= \max_t \left\| D^{-1}Q\beta_{F,t} + (1-q)\hat{\phi}_{F,t}^{(\ell-2)} + O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})\right\| \\
&\leq \left\|D^{-1}Q\right\| \max_t \left\|\beta_{F,t}\right\| + (1-q)\max_t \left\|\hat{\phi}_{F,t}^{(\ell-2)}\right\| + O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4}) \\
&= O_P((N/\ln N)^{-1/2}) + O_P((N/\ln N)^{-1/2}) + o_P((N \vee T)^{-1/2}) = O_P((N/\ln N)^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
\max_i \left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\| &= \max_i \left\| (Q')^{-1}\beta_{\Lambda,i} + (1-q)\hat{\phi}_{\Lambda,i}^{(\ell-2)} + O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N)\right\| \\
&\leq \left\|(Q')^{-1}\right\| \max_i \left\|\beta_{\Lambda,i}\right\| + (1-q)\max_i \left\|\hat{\phi}_{\Lambda,i}^{(\ell-2)}\right\| + O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N) \\
&= O_P((T/\ln T)^{-1/2}) + O_P((T/\ln T)^{-1/2}) + o_P((N \vee T)^{-1/2}) = O_P((T/\ln T)^{-1/2}).
\end{aligned}$$

(ii) By the decomposition in (A.8), $\hat{r}_{F,t}^{(\ell-1)} = \hat{F}_t^{(\ell-1)} - \hat{H}^{(\ell-1)\prime}F_t^{\ell-1} - \hat{\phi}_{F,t}^{(\ell-1)} = \hat{a}_{1t}^{(\ell-1)} + \hat{a}_{3t}^{(\ell-1)} + (\hat{a}_{2t}^{(\ell-1)} - \hat{\phi}_{F,t}^{(\ell-1)})$. Following the proof of Theorem 2.4(i) and using Assumption A.5 and the fact that $\max_t \left\|F_t^0\right\| = O_P(T^{\gamma_1/4})$, it is easy to show that

$$\max_t \left\|\hat{a}_{1t}^{(\ell-1)}\right\| = O_P(T^{-1/2}\delta_{NT}^{-1} + T^{-1+\gamma_1/4}), \quad \max_t \left\|\hat{a}_{3t}^{(\ell-1)}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}),$$

and $\max_t \left\|\hat{a}_{2t}^{(\ell-1)} - \hat{\phi}_{F,t}^{(\ell-1)}\right\| = O_P(\delta_{NT}^{-2}\ln N)$. It follows that $\max_t \left\|\hat{r}_{F,t}^{(\ell-1)}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T + T^{-1+3\gamma_1/4})$. For $\hat{r}_{\Lambda,i}^{(\ell-1)}$, we have

$$\hat{r}_{\Lambda,i}^{(\ell-1)} = \hat{\lambda}_i^{(\ell-1)} - (\hat{H}^{(\ell-1)})^{-1}\lambda_i^0 - \hat{\phi}_{\Lambda,i}^{(\ell-1)} = \hat{B}_{2i}^{(\ell-1)} + \hat{B}_{3i}^{(\ell-1)} + (\hat{B}_{1i}^{(\ell-1)} - \hat{\phi}_{\Lambda,i}^{(\ell-1)}),$$

where $\hat{B}_{li}^{(\ell)}$'s are defined in the proof of Theorem 2.4(ii). Following the proof of Theorem 2.4(ii) and using the fact that $\max_i \|\lambda_i^0\| = O_P(N^{\gamma_2/4})$, $\max_i \frac{1}{T}\sum_{t=1}^T \varepsilon_{it}^2 = O_P(1)$, and $\tilde{q} - q = O_P((NT)^{-1/2})$ we have by Lemma A.4

$$\max_i \left\|\hat{B}_{2i}^{(\ell-1)}\right\| = O_P(N^{-1/2+\gamma_2/4}\delta_{NT}^{-1} + \delta_{NT}^{-2}\ln N), \ \ \max_i \left\|\hat{B}_{3i}^{(\ell-1)}\right\| = O_P(N^{\gamma_2/4}\delta_{NT}^{-2}),$$

and $\max_i \left\|\hat{B}_{1i}^{(\ell-1)} - \hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\| = O_P(\delta_{NT}^{-2}\ln N)$. It follows that $\max_i \left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\| = O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N)$.

(iii) The proof is similar to the $\ell = 1$ case by replacing the superscript 0 by $\ell - 1$ throughout the proof.

(iv) By Assumption A.5 and Lemma A.3(x) below

$$\begin{aligned}
\left(\hat{H}^{(\ell-1)\prime}\right)^{-1}\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(\ell-1)}\varepsilon_{it}g_{it} &= \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\left[\varepsilon_{is}g_{is} + \eta_{is}^{(\ell-1)}\bar{g}_{is}\right]\varepsilon_{it}g_{it} \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0 E\left(\varepsilon_{is}\varepsilon_{it}\right)g_{is}g_{it} + \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\left[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})\right]g_{is}g_{it} \\
&\quad + \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\eta_{is}^{(\ell-1)}\bar{g}_{is}\varepsilon_{it}g_{it} \\
&= O_P(T^{-1+\gamma_1/4}) + O_P\left(\delta_{NT}^{-2}\ln N\right) + O_P(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N).
\end{aligned}$$

Then $\max_i \left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(\ell-1)}\varepsilon_{it}g_{it}\right\| = O_P\left(T^{-1+\gamma_1/4} + \delta_{NT}^{-2}\ln N\right)$.

Note that $\left\|\frac{1}{N}\sum_{i=1}^N \hat{\phi}_{\Lambda,i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\| \leq \left\|\hat{H}^{(\ell-1)\prime}\right\|\left\|\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\varepsilon_{is}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\|$. Using $\bar{g}_{it} = (1-q) - (g_{it} - q)$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\varepsilon_{is}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it} &= \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\lambda_i^{0\prime}\left[\varepsilon_{is}g_{is} + \eta_{is}^{(\ell-1)}\bar{g}_{is}\right]\bar{g}_{it} \\
&= \frac{1-q}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\lambda_i^{0\prime}\varepsilon_{is}g_{is} + \frac{1-q}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is} \\
&\quad - \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\lambda_i^{0\prime}\varepsilon_{is}g_{is}(g_{it} - q) - \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}(g_{it} - q).
\end{aligned}$$

It is easy to show that the first term is $O_P\left(\delta_{NT}^{-2}\right)$ by Chebyshev inequality. The second term is $O_P(\delta_{NT}^{-2})$ by Lemma A.3(viii) below. The third term is $O_P(\delta_{NT}^{-2}\ln N)$ by Assumption A.5(iii). By Lemma A.3(iii),

$$\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is}(g_{it} - q) = \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T F_s^0\lambda_i^{0\prime}\eta_{1,is}^{(\ell-1)}\bar{g}_{is}(g_{it} - q) + O_P(\delta_{NT}^{-2}\ln N)$$

uniformly in t. Now we make the following decomposition

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\eta_{1,is}^{(\ell-1)}\bar{g}_{is}(g_{it}-q) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}[F_s^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,s}^{(\ell-1)} \\
&\quad +\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,s}^{(\ell-1)}+F_s^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}]\bar{g}_{is}(g_{it}-q) \\
&\equiv II_{1t}+II_{2t}+II_{3t}+II_{4t}.
\end{aligned}
$$

For II_{3t} and II_{4t}, we apply Lemma A.3(ii) to obtain the rough bound

$$
\begin{aligned}
\max_t\|II_{3t}\| &\leq \max_t\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,s}^{(\ell-1)}\bar{g}_{is}(g_{it}-q)\right\| \\
&\leq O_P(1)\max_s\left\|\hat{r}_{F,s}^{(\ell-1)}\right\| = O_P(T^{\gamma_1/4}\delta_{NT}^{-2}\ln T+T^{-1+3\gamma_1/4}) \text{ and} \\
\max_t\|II_{4t}\| &= \max_t\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\left[F_s^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\right]\bar{g}_{is}(g_{it}-q)\right\| \\
&\leq O_P(1)\max_i\left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\| = O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N).
\end{aligned}
$$

By Lemma A.5(i), we have

$$
\begin{aligned}
\max_t\|II_{1t}\| &= \max_t\left\|\frac{1}{NT}\sum_{s=1}^{T}\left[\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}F_s^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right]\bar{g}_{is}(g_{it}-q)\right\| \\
&\leq \max_t\left\|\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\text{tr}\left[\lambda_i^{0\prime}F_s^{0\prime}\hat{H}^{(\ell-1)}(Q^{\prime})^{-1}\beta_{\Lambda,i}\bar{g}_{is}(g_{it}-q)\right]\right\| \\
&\quad +\max_t\left\|\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\text{tr}\left[\lambda_i^{0\prime}F_s^{0\prime}\hat{H}^{(\ell-1)}(1-q)\hat{\phi}_{\Lambda,i}^{(\ell-2)}\bar{g}_{is}(g_{it}-q)\right]\right\|+O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N) \\
&\leq O_P(1)\max_{s,t}\left\|\frac{1}{N}\sum_{i=1}^{N}\beta_{\Lambda,i}\lambda_i^{0\prime}\bar{g}_{is}(g_{it}-q)\right\|+O_P(1)\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-2)}\lambda_i^{0\prime}\bar{g}_{is}(g_{it}-q)\right\| \\
&\quad +O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N) \\
&= O_P(\delta_{NT}^{-2}\ln N)+O_P(T^{-1+\gamma_1/2}+\delta_{NT}^{-2}\ln N)+O_P(N^{\gamma_2/4}\delta_{NT}^{-2}\ln N) \\
&= O_P(T^{-1+\gamma_1/2}+N^{\gamma_2/4}\delta_{NT}^{-2}\ln N)
\end{aligned}
$$

Similarly, using Lemma A.5(ii), we can show that

$$
\max_t\|II_{2t}\| = \max_t\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}F_s^0\lambda_i^{0\prime}\left[\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,s}^{(\ell-1)}\right]\bar{g}_{is}(g_{it}-q)\right\| = O_P(T^{-1+\gamma_1/2}+\delta_{NT}^{-2}\ln N).
$$

Noting that $\hat{r}_{\Lambda,i}^{(\ell-1)}=\hat{\lambda}_i^{(\ell-1)}-(\hat{H}^{(\ell-1)})^{-1}\lambda_i^0-\hat{\phi}_{\Lambda,i}^{(\ell-1)}=\hat{B}_{2i}^{(\ell-1)}+\hat{B}_{3i}^{(\ell-1)}$, we have $\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{r}_{\Lambda,i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\|$ $\leq\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}[\hat{B}_{2i}^{(\ell-1)}+\hat{B}_{3i}^{(\ell-1)}]\lambda_i^{0\prime}\bar{g}_{it}\right\|$, where $\hat{B}_{li}^{(\ell-1)}$'s are defined in the proof of Theorem 2.4(ii).

By Lemma A.4(iv)-(v), we have

$$\begin{aligned}\max_t \left\| \frac{1}{N}\sum_{i=1}^N \hat{B}_{2i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\| &= \max_t \frac{1}{NT}\left\|(\hat{H}^{(\ell-1)})^{-1}\sum_{i=1}^N\sum_{s=1}^T(\hat{H}^{(\ell-1)\prime}F_s^0-\hat{F}_s^{(\ell-1)})\varepsilon_{is}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\| \\ &\leq O_P(1)\max_s\left\|\frac{1}{T}\sum_{s=1}^T(\hat{H}^{(\ell-1)\prime}F_s^0-\hat{F}_s^{(\ell-1)})\varepsilon_{is}^{(\ell-1)}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right),\end{aligned}$$

and

$$\begin{aligned}\max_t \left\| \frac{1}{N}\sum_{i=1}^N \hat{B}_{3i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\| &= \max_t \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T\hat{F}_s^{(\ell-1)}(\hat{H}^{(\ell-1)\prime}F_s^0-\hat{F}_s^{(\ell-1)})'(\hat{H}^{(\ell-1)\prime})^{-1}\lambda_i^{0\prime}\bar{g}_{it}\right\| \\ &\leq O_P(1)\left\|\frac{1}{N}\sum_{s=1}^T\hat{F}_s^{(\ell-1)}(\hat{H}^{(\ell-1)\prime}F_s^0-\hat{F}_s^{(\ell-1)})'\right\| = O_P(\delta_{NT}^{-2}).\end{aligned}$$

In sum, we have $\max_t \left\|\hat{H}^{(\ell-1)\prime}\frac{1}{N}\sum_{i=1}^N \hat{r}_{\Lambda,i}^{(\ell-1)}\lambda_i^{0\prime}\bar{g}_{it}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$.

(v) By the definition of $\hat{\phi}_{F,t}^{(\ell-1)}$ and $\varepsilon_{jt}^{(\ell-1)}$, we have

$$\begin{aligned}&\max_i\left\|\frac{1}{T}\sum_{t=1}^T\hat{\phi}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| \\ \leq\; & O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0\varepsilon_{jt}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| \\ \leq\; & O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\varepsilon_{jt}g_{jt}\bar{g}_{it}\right\| + O_P(1)\max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\eta_{jt}^{(\ell-2)}\bar{g}_{jt}\bar{g}_{it}\right\|.\end{aligned}$$

We can show that the first term is $O_P(\delta_{NT}^{-2}\ln N)$ by applying Assumption A.5(iii). For the second term, we have by Lemma A.3(viii) and (iii)

$$\begin{aligned}&\max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\eta_{jt}^{(\ell-2)}\bar{g}_{jt}\bar{g}_{it}\right\| \\ \leq\; & \left\|\frac{q-1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\eta_{jt}^{(\ell-2)}\bar{g}_{jt}\right\| + \max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\eta_{jt}^{(\ell-2)}\bar{g}_{jt}(g_{it}-q)\right\| \\ =\; & O_P\left(\delta_{NT}^{-2}\right) + \max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\eta_{jt}^{(\ell-2)}\bar{g}_{jt}(g_{it}-q)\right\| \\ =\; & \max_i\left\|\frac{1}{NT}\sum_{t=1}^T\sum_{j=1}^N\lambda_j^0F_t^{0\prime}\eta_{1,jt}^{(\ell-2)}\bar{g}_{jt}(g_{it}-q)\right\| + O_P\left(\delta_{NT}^{-2}\ln N\right).\end{aligned}$$

Noting that $\eta^{(\ell)}_{1,it} = F^{0\prime}_t \hat{H}^{(\ell)} \hat{\phi}^{(\ell)}_{\Lambda,i} + \lambda^{0\prime}_i (\hat{H}^{(\ell)\prime})^{-1} \hat{\phi}^{(\ell)}_{F,t} + \lambda^{0\prime}_i (\hat{H}^{(\ell)\prime})^{-1} \hat{r}^{(\ell)}_{F,t} + F^{0\prime}_t \hat{H}^{(\ell)\prime} \hat{r}^{(\ell)}_{\Lambda,i}$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\lambda_j^0 F_t^{0\prime}\eta_{1,jt}^{(\ell-2)}\bar{g}_{jt}(g_{it}-q) &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\lambda_j^0 F_t^{0\prime}[F_t^{0\prime}\hat{H}^{(\ell-2)}\hat{\phi}_{\Lambda,j}^{(\ell-2)} + \lambda_j^{0\prime}(\hat{H}^{(\ell-2)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-2)} \\
&\quad + \lambda_j^{0\prime}(\hat{H}^{(\ell-2)\prime})^{-1}\hat{r}_{F,t}^{(\ell-2)} + F_t^{0\prime}\hat{H}^{(\ell-2)\prime}\hat{r}_{\Lambda,j}^{(\ell-2)}]\bar{g}_{jt}(g_{it}-q) \\
&\equiv III_{1i} + III_{2i} + III_{3i} + III_{4i}.
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
\max_i \|III_{1i}\| &= \max_i \left\| \frac{1}{N}\sum_{j=1}^{N}\lambda_j^0 \hat{\phi}_{\Lambda,j}^{(\ell-2)\prime}\hat{H}^{(\ell-2)\prime}\frac{1}{T}\sum_{t=1}^{T}F_t^0 F_t^{0\prime}\bar{g}_{jt}(g_{it}-q)\right\| \\
&\le \max_j \left\|\hat{\phi}_{\Lambda,j}^{(\ell-2)}\right\| \max_j \left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0 F_t^{0\prime}\bar{g}_{jt}(g_{it}-q)\right\| \\
&= O_P((T/\ln T)^{-1/2})O_P((T/\ln T)^{-1/2} + T^{-1+\gamma_1/2}) = O_P\left(\delta_{NT}^{-2}\ln N\right).
\end{aligned}$$

Similarly, we can show that $\max_i \|III_{4i}\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$ and $\max_i \|III_{4i}\| = O_P\left(\delta_{NT}^{-2}\ln N + N^{-1+\gamma_2/2}\right)$ for $l = 2,3$. Then $\max_i \left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\phi}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| = O_P\left(\delta_{NT}^{-2}\ln N + N^{-1+\gamma_2/2}\right)$.

Noting that $\hat{r}_{F,t}^{(\ell-1)} = \hat{F}_t^{(\ell-1)} - \hat{H}^{(\ell-1)\prime}F_t^0 - \hat{\phi}_{F,t}^{(\ell-1)} = \hat{a}_{1t}^{(\ell-1)} + \hat{a}_{3t}^{(\ell-1)}$ by (A.8), we have

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{r}_{F,t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| \le \left\|\frac{1}{T}\sum_{t=1}^{T}\hat{a}_{1t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T}\hat{a}_{3t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\|.$$

Note that

$$\begin{aligned}
\max_i \left\|\frac{1}{T}\sum_{t=1}^{T}\hat{a}_{1t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| &\le O_P(1)\max_i \left\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^{(\ell-1)}\sum_{j=1}^{N}\varepsilon_{jt}^{(\ell-1)}\varepsilon_{js}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| \\
&= O_P(1)\max_i \left\|\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}F_s^0\varepsilon_{jt}^{(\ell-1)}\varepsilon_{js}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) \\
&\le O_P(1)\max_i \frac{1}{N}\sum_{j=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0\varepsilon_{jt}^{(\ell-1)}\bar{g}_{it}\right\|\left\|\frac{1}{T}\sum_{s=1}^{T}F_s^0\varepsilon_{js}^{(\ell-1)}\right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) \\
&\le O_P(1)\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}F_s^0\varepsilon_{js}^{(\ell-1)}\right\|^2 + O_P\left(\delta_{NT}^{-2}\ln N\right).
\end{aligned}$$

Using the decomposition $\varepsilon_{jt}^{(\ell-1)} = \varepsilon_{jt}g_{jt} + \eta_{jt}^{(\ell-2)}\bar{g}_{jt}$ and Assumption A.5, we can show that $\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N}\varepsilon_{jt}^{(\ell-1)}\varepsilon_{js}^{(\ell-1)}F_s^0 F_t^{0\prime}\bar{g}_{it} = O_P\left(\delta_{NT}^{-2}\ln N\right)$. Then $\max_i \left\|\frac{1}{T}\sum_{t=1}^{T}\hat{a}_{1t}^{(\ell-1)}F_t^{0\prime}\bar{g}_{it}\right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$.

Similarly,

$$\begin{aligned}
\max_i \left\| \frac{1}{T}\sum_{t=1}^T \hat{a}_{3t}^{(\ell)} F_t^{0\prime}\bar{g}_{it} \right\| &\le O_P(1)\max_i \left\| \frac{1}{T}\sum_{t=1}^T \left[\frac{1}{NT}\sum_{s=1}^T \hat{F}_s^{(\ell)} \sum_{j=1}^N \lambda_j^{0\prime} F_t^0 \varepsilon_{js}^{(\ell)} \right] F_t^{0\prime}\bar{g}_{it} \right\| \\
&= O_P(1)\max_i \left\| \frac{1}{T}\sum_{t=1}^T \left(\frac{1}{NT}\sum_{s=1}^T\sum_{j=1}^N F_s^0 \varepsilon_{js}^{(\ell)} \lambda_j^{0\prime} \right) F_t^0 F_t^{0\prime}\bar{g}_{it} \right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) \\
&= O_P(1) \left\| \frac{1}{NT}\sum_{s=1}^T\sum_{i=1}^N F_j^0 \varepsilon_{is}^{(\ell)} \lambda_j^{0\prime} \right\| + O_P\left(\delta_{NT}^{-2}\ln N\right) = O_P\left(\delta_{NT}^{-2}\ln N\right).
\end{aligned}$$

It follows that $\max_i \left\| \frac{1}{T}\sum_{t=1}^T \hat{r}_{F,t}^{(\ell-1)} F_t^{0\prime}\bar{g}_{it} \right\| = O_P\left(\delta_{NT}^{-2}\ln N\right)$.

(vi) Note that $\frac{1}{N}\sum_{i=1}^N \left\|\eta_{it}^{(\ell-1)}\right\|^2 \le \frac{2}{N}\sum_{i=1}^N \left\|\eta_{1,it}^{(\ell-1)}\right\|^2 + \frac{2}{N}\sum_{i=1}^N \left\|\eta_{2,it}^{(\ell-1)}\right\|^2$, where the second term is bounded above by $O_P(\delta_{NT}^{-4}(\ln N)^2)$ by (iii). For the first term, we have

$$\begin{aligned}
&\max_t \frac{1}{N}\sum_{i=1}^N \left\|\eta_{1,it}^{(\ell-1)}\right\|^2 \\
\le\ & \max_t \frac{1}{N}\sum_{i=1}^N \left\| F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)} + F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)} \right\|^2 \\
\le\ & 4\left\|\hat{H}^{(\ell-1)}\right\| \max_t \left\|F_t^0\right\|^2 \frac{1}{N}\sum_{i=1}^N \left(\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2 + \left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\|^2 \right) \\
& + 4\left\|(\hat{H}^{(\ell-1)\prime})^{-1}\right\| \left\{ \max_t \left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\|^2 + \max_t \left\|\hat{r}_{F,t}^{(\ell-1)}\right\|^2 \right\} \frac{1}{N}\sum_{i=1}^N \max_i \left\|\lambda_i^0\right\|^2 \\
=\ & O_P(T^{-1+\gamma_1/2} + N^{-1}\ln N).
\end{aligned}$$

It follows that $\frac{1}{N}\sum_{i=1}^N \left\|\eta_{it}^{(\ell-1)}\right\|^2 = O_P(T^{-1+\gamma_1/2} + N^{-1}\ln N)$. Similarly, we can show that $\max_t \frac{1}{T}\sum_{t=1}^T \left\|\eta_{it}^{(\ell-1)}\right\|^2 = O_P(N^{-1+\gamma_2/2} + T^{-1}\ln N)$.

(vii) Recall that $\kappa_t = 1 + \left\|F_t^0\right\|^2$. By the CS inequality and (iii), $\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \kappa_t(\eta_{it}^{(\ell-1)})^2 \le \frac{2}{NT}\sum_{t=1}^T\sum_{i=1}^N \kappa_t(\eta_{1,it}^{(\ell-1)})^2 + O_P(\delta_{NT}^{-4}(\ln N)^2)$. Using $\eta_{1,it}^{(\ell)} = F_t^{0\prime}\hat{H}^{(\ell)}\hat{\phi}_{\Lambda,i}^{(\ell)} + \lambda_i^{0\prime}(\hat{H}^{(\ell)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell)} + \lambda_i^{0\prime}(\hat{H}^{(\ell)\prime})^{-1}\hat{r}_{F,t}^{(\ell)} + F_t^{0\prime}\hat{H}^{(\ell)\prime}\hat{r}_{\Lambda,i}^{(\ell)} = \sum_{l=1}^4 \eta_{1,it}^{(\ell)}(l)$, we have

$$\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \kappa_t(\eta_{1,it}^{(\ell-1)})^2 \le 4\sum_{l=1}^4 \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \kappa_t\left[\eta_{1,it}^{(\ell-1)}(l)\right]^2 \equiv 4\sum_{l=1}^4 II_{2,l}.$$

Noting that $\frac{1}{N}\sum_{i=1}^N \left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2 = O_P\left(T^{-1}\right)$, we can readily show $II_{2,1} \le \left\|\hat{H}^{(\ell-1)}\right\|^2 \frac{1}{T}\sum_{t=1}^T \left\|F_t^0\right\|^4 \frac{1}{N}\sum_{i=1}^N$

$\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^2 = O_P\left(T^{-1}\right)$. By Lemma A.5(i)

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\|^2 &\leq \frac{2}{T}\sum_{t=1}^{T}\kappa_t\left\|D^{-1}Q\beta_{F,t}+(1-q)\,\hat{\phi}_{F,t}^{(\ell-2)}\right\|^2 + O_P\left(T^{\gamma_1/2}\delta_{NT}^{-4}(\ln T)^2 + T^{-2+3\gamma_1/2}\right) \\
&\leq O_P(1)\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\beta_{F,t}\right\|^2 + O_P(1)\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\hat{\phi}_{F,t}^{(\ell-2)}\right\|^2 + O_P\left(\delta_{NT}^{-2}\right) \\
&= O_P\left(\delta_{NT}^{-2}\right),
\end{aligned}$$

we have $II_{2,2} \leq \left\|(\hat{H}^{(\ell-1)\prime})^{-1}\right\|^2 \frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^{0\prime}\right\|^2 \frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|\hat{\phi}_{F,t}^{(\ell-1)}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. Similarly, we have

$$II_{2,3} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t\left|\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}\right|^2 \leq O_P(1)\max_t\left\|\hat{r}_{F,t}^{(\ell-1)}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\kappa_t\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^{0\prime}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right),$$

and

$$II_{2,4} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t\left|F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}\right|^2 \leq O_P(1)\max_i\left\|\hat{r}_{\Lambda,i}^{(\ell-1)}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\kappa_t\left\|F_t^0\right\|^2 = O_P\left(\delta_{NT}^{-2}\right).$$

It follows that $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\kappa_t(\eta_{it}^{(\ell-1)})^2 = O_P\left(\delta_{NT}^{-2}\right)$.

(viii) Note that $\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\eta_{is}^{(\ell-1)}\bar{g}_{is} = \sum_{l=1}^{2}\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0\lambda_i^{0\prime}\eta_{l,is}^{(\ell-1)}\bar{g}_{is} \equiv \sum_{l=1}^{2}II_{3,l}$. We only show $II_{3,1} = O_P\left(\delta_{NT}^{-2}\right)$ as the other term is of smaller order. Note that

$$\begin{aligned}
II_{3,1} &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i^{0\prime}[F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)} + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)} \\
&\quad + \lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)} + F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}]\bar{g}_{it} \\
&\equiv II_{3,1}(1) + II_{3,1}(2) + II_{3,1}(3) + II_{3,1}(4).
\end{aligned}$$

Let λ_{il}^0 and F_{sl}^0 denote the lth element of λ_i^0 and F_s^0, respectively. Let $II_{3,1lr}(\cdot)$ denote the (l,r)th element of $B_{3,1}(\cdot)$. Noting that $\bar{g}_{is} = (1-q)+(q-g_{is})$, we have

$$\begin{aligned}
&\left\|II_{3,1lr}(1)\right\| \\
&= \left\|\frac{1}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(\ell-1)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\bar{g}_{it}\lambda_{il}^0\right\| \\
&\leq \left\|\frac{1-q}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(\ell-1)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\lambda_{il}^0\right\| + \left\|\frac{1}{NT}\sum_{t=1}^{T}F_{tr}^0F_t^{0\prime}\hat{H}^{(\ell-1)}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}(g_{it}-q)\lambda_{il}^0\right\| \\
&\equiv II_{3,1lr}(1,1) + II_{3,1lr}(1,2).
\end{aligned}$$

For $II_{3,1lr}(1,1)$, we have

$$\begin{aligned}
II_{3,1lr}(1,1) &\leq O_P(1)\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\lambda_{il}^0\right\| = O_P(1)\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{H}^{(\ell-1)\prime}[\beta_{\Lambda,i}+(1-q)\,\hat{\phi}_{\Lambda,i}^{(\ell-2)}]\lambda_{il}^0\right\| \\
&\leq O_P(1)\left\{\frac{1}{N}\sum_{i=1}^{N}\beta_{\Lambda,i}\lambda_{il}^0 + (1-q)\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-2)}\lambda_{il}^0\right\} = O_P(\delta_{NT}^{-2}).
\end{aligned}$$

For $II_{3,1lr}(1,2)$, we have

$$\begin{aligned} II_{3,1lr}(1,2) &= \left\| \frac{1}{N}\sum_{i=1}^{N}\lambda_{il}^{0}\hat{\phi}_{\Lambda,i}^{(\ell-1)\prime}\hat{H}^{(\ell-1)\prime}\left[\frac{1}{T}\sum_{t=1}^{T}F_{t}^{0}F_{tr}^{0}(g_{it}-q)\right]\right\| \\ &\leq \left\|\hat{H}^{(\ell-1)}\right\|\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_{i}^{0}\right\|^{2}\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^{2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}F_{t}^{0}F_{tr}^{0}(g_{it}-q)\right\|^{2}\right\}^{1/2} \\ &= O_P\left(\delta_{NT}^{-1}\right)O_P(T^{-1/2}) \end{aligned}$$

as we can show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_{i}^{0}\right\|^{2}\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|^{2}=O_P\left(\delta_{NT}^{-2}\right)$ and $\frac{1}{N}\sum_{i=1}^{N}E\left\|\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}F_{sr}^{0}(g_{is}-q)\right\|^{2}=O\left(T^{-1}\right)$. Then $II_{3,1}(1)=O_P\left(\delta_{NT}^{-2}\right)$. Similarly,

$$\begin{aligned} &\left\|II_{3,1rl}(2)\right\| \\ =\;& \left\|\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\lambda_{ir}^{0}\lambda_{i}^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,s}^{(\ell-1)}F_{sl}^{0}\bar{g}_{is}\right\| \\ \leq\;& \left\|\frac{1-q}{N}\sum_{i=1}^{N}\lambda_{ir}^{0}\lambda_{i}^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\frac{1}{T}\sum_{s=1}^{T}\hat{\phi}_{F,s}^{(\ell-1)}F_{sl}^{0}\right\|+\left\|\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}\lambda_{ir}^{0}\lambda_{i}^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,s}^{(\ell-1)}F_{sl}^{0}\left(g_{is}-q\right)\right\| \\ \equiv\;& II_{3,1rl}(2,1)+II_{3,1rl}(2,2). \end{aligned}$$

For $II_{3,1rl}(2,1)$, we have $II_{3,1rl}(2,1)\leq O_P(1)\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{\phi}_{F,s}^{(\ell-1)}F_{sl}^{0}\right\|=O_P\left(\delta_{NT}^{-2}\right)$ as we can show that $\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{\phi}_{F,s}^{(\ell-1)}F_{sl}^{0}\right\|=O_P\left(\delta_{NT}^{-2}\right)$ by following the analysis of $\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\lambda_{il}^{0}\right\|$. For $II_{3,1rl}(2,2)$, we have

$$\begin{aligned} II_{3,1rl}(2,2) &= \left\|\frac{1}{T}\sum_{s=1}^{T}F_{sl}^{0}\hat{\phi}_{F,s}^{(\ell-1)\prime}(\hat{H}^{(\ell-1)})^{-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_{i}^{0}\lambda_{ir}^{0}\left(g_{is}-q\right)\right\| \\ &\leq \left\|(\hat{H}^{(\ell-1)})^{-1}\right\|\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|F_{s}^{0}\right\|^{2}\left\|\hat{\phi}_{F,s}^{(\ell-1)}\right\|^{2}\right\}^{1/2}\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{i}^{0}\lambda_{ir}^{0}\left(g_{is}-q\right)\right\|^{2}\right\}^{1/2} \\ &= O_P\left(\delta_{NT}^{-1}\right)O_P(N^{-1/2}). \end{aligned}$$

So $II_{3,1}(2)=O_P\left(\delta_{NT}^{-2}\right)$. Analogously, we can show that $II_{3,1}(l)=O_P\left(\delta_{NT}^{-2}\right)$ for $l=3,4$. Then $II_{3,1}=O_P\left(\delta_{NT}^{-2}\right)$.

(ix) By (vi) and the fact that $\frac{1}{N}\sum_{i=1}^{N}E\left(\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}\varepsilon_{is}g_{is}\right)^{2}=O\left(T^{-1}\right)$,

$$\begin{aligned} \max_{t}\left\|\frac{1}{NT}\sum_{s=1}^{T}F_{s}^{0}\sum_{i=1}^{N}\eta_{it}^{(\ell-1)}\bar{g}_{it}\varepsilon_{is}g_{is}\right\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\eta_{it}^{(\ell-1)}\bar{g}_{it}\left(\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}\varepsilon_{is}g_{is}\right)\right\| \\ &\leq \left\{\max_{t}\frac{1}{N}\sum_{i=1}^{N}(\eta_{it}^{(\ell-1)})^{2}\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}F_{s}^{0}\varepsilon_{is}g_{is}\right)^{2}\right\}^{1/2} \\ &= O_P\left((T^{-1/2+\gamma_1/4}+(N/\ln N)^{-1/2}\right)O_P\left(T^{-1/2}\right) \\ &= O_P\left((T^{-1+\gamma_1/4}+(NT/\ln N)^{-1/2}\right). \end{aligned}$$

(x) Note that $\frac{1}{NT}\sum_{s=1}^{T}F_s^0\sum_{i=1}^{N}\varepsilon_{it}g_{it}\eta_{is}^{(\ell-1)}\bar{g}_{is}\leq\sum_{l=1}^{2}\frac{1}{NT}\sum_{s=1}^{T}F_s^0\sum_{i=1}^{N}\varepsilon_{it}g_{it}\eta_{l,is}^{(\ell-1)}\bar{g}_{is}\equiv\sum_{l=1}^{2}II_{4,lt}$. One can bound $II_{4,2t}$ by $O_P\left(\delta_{NT}^{-2}\ln N\right)$ by using the uniform bound for $\eta_{2,is}^{(\ell-1)}$ in (iii). For $II_{4,1t}$, we have

$$
\begin{aligned}
II_{4,1t} &= \frac{1}{NT}\sum_{s=1}^{T}F_s^0\sum_{i=1}^{N}\varepsilon_{it}g_{it}[F_t^{0\prime}\hat{H}^{(\ell-1)}\hat{\phi}_{\Lambda,i}^{(\ell-1)}+\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{\phi}_{F,t}^{(\ell-1)} \\
&\quad +\lambda_i^{0\prime}(\hat{H}^{(\ell-1)\prime})^{-1}\hat{r}_{F,t}^{(\ell-1)}+F_t^{0\prime}\hat{H}^{(\ell-1)\prime}\hat{r}_{\Lambda,i}^{(\ell-1)}]\bar{g}_{is} \\
&\equiv II_{4,1t}(1)+II_{4,1t}(2)+II_{4,1t}(3)+II_{4,1t}(4).
\end{aligned}
$$

For $II_{4,1t}(1)$, by (i), (iv) and the fact $\frac{1}{N}\sum_{i=1}^{N}E\left\|\frac{1}{T}\sum_{s=1}^{T}F_s^0F_s^{0\prime}(g_{is}-q)\right\|^2=O\left(T^{-1}\right)$, we have

$$
\begin{aligned}
\max_t II_{4,1t}(1) &\leq \max_t\left\|\frac{1-q}{T}\sum_{s=1}^{T}F_s^0F_s^{0\prime}\hat{H}^{(\ell-1)}\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\varepsilon_{it}g_{it}\right\| \\
&\quad +\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}g_{it}\hat{\phi}_{\Lambda,i}^{(\ell-1)\prime}\hat{H}^{(\ell-1)\prime}\frac{1}{T}\sum_{s=1}^{T}F_s^0F_s^{0\prime}(g_{is}-q)\right\| \\
&\leq O_P(1)\max_t\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{\phi}_{\Lambda,i}^{(\ell-1)}\varepsilon_{it}g_{it}\right\| \\
&\quad +\max_i\left\|\hat{\phi}_{\Lambda,i}^{(\ell-1)}\right\|\left\{\max_t\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}F_s^0F_s^{0\prime}(g_{is}-q)\right\|^2\right\}^{1/2} \\
&= O_P(T^{-1+\gamma_1/2}+\delta_{NT}^{-2}\ln N)+O_P((T/\ln T)^{-1/2})O_P(T^{-1/2})=O_P(T^{-1+\gamma_1/2}+\delta_{NT}^{-2}\ln N).
\end{aligned}
$$

For $II_{3,1t}(2)$, we have

$$
\begin{aligned}
\max_t II_{4,1t}(2) &= \left\|\frac{1}{T}\sum_{s=1}^{T}F_s^0\hat{\phi}_{F,s}^{(\ell-1)\prime}[\hat{H}^{(\ell-1)}]^{-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}\bar{g}_{is}\right\| \\
&\leq \left\|\frac{1-q}{T}\sum_{s=1}^{T}F_s^0\hat{\phi}_{F,s}^{(\ell-1)\prime}[\hat{H}^{(\ell-1)}]^{-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}\right\| \\
&\quad +\left\|\frac{1}{T}\sum_{s=1}^{T}F_s^0\hat{\phi}_{F,s}^{(\ell-1)\prime}[\hat{H}^{(\ell-1)}]^{-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}(g_{is}-q)\right\| \\
&\leq O_P(1)\max_s\left\|\hat{\phi}_{F,s}^{(\ell-1)}\right\|\max_s\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}\bar{g}_{is}\right\| \\
&\quad +O_P(1)\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|F_s^0\hat{\phi}_{F,s}^{(\ell-1)\prime}\right\|^2\right\}^{1/2}\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\varepsilon_{it}g_{it}(g_{is}-q)\right\|^2\right\}^{1/2} \\
&= O_P((N/\ln N)^{-1/2})O_P((N/\ln N)^{-1/2})+O_P((N/\ln N)^{-1/2})O_P(\delta_{NT}^{-1})=O_P(\delta_{NT}^{-2}\ln N).
\end{aligned}
$$

Analogously, we can show that $\max_t\|II_{4,1t}(l)\|=O_P(\delta_{NT}^{-2}\ln N)$ for $l=3,4$. Then $\max_t\|II_{3,1t}\|=O_P\left(T^{-1+\gamma_1/2}+\delta_{NT}^{-2}\ln N\right)$. ■

Proof of Lemma A.6. (i) $\max_i \frac{1}{T}\sum_{t=1}^{T}|\hat{\varepsilon}_{it}-\varepsilon_{it}|^2 = O_P\left(\underline{m}^{-1}\ln T\right)$. Noting that $\hat{\varepsilon}_{it}-\varepsilon_{it} = \hat{\lambda}_i'\hat{F}_t^{(0)} - \lambda_i^{0\prime}F_t^{(0)} = (\hat{\lambda}_i^{(0)} - \hat{H}^{-1}\lambda_i^0)'\hat{F}_t + \lambda_i^{0\prime}\{[\hat{H}']^{-1}\hat{F}_t - F_t^0\}$, we have

$$\begin{aligned}
\max_i \frac{1}{T}\sum_{t=1}^{T}|\hat{\varepsilon}_{it}-\varepsilon_{it}|^2 &= \max_i \frac{1}{T}\sum_{t=1}^{T}\left|(\hat{\lambda}_i - \hat{H}^{-1}\lambda_i^0)'\hat{F}_t + \lambda_i^{0\prime}\{(\hat{H}')^{-1}\hat{F}_t - F_t^0\}\right|^2 \\
&\leq 2R\max_i\left\|\hat{\lambda}_i - \hat{H}^{-1}\lambda_i^0\right\|^2 + 2\max_i\left\|\lambda_i^0\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - \hat{H}'F_t^0\right\|^2\left\|(\hat{H}')^{-1}\right\|^2 \\
&= O_P(T^{-1}\ln T) + O_P(N^{\gamma_2/2})O_P(N^{-1}) = O_P(N^{-1+\gamma_2/2} + T^{-1}\ln T).
\end{aligned}$$

(ii) Note that

$$\begin{aligned}
\max_{i,t}|\hat{\varepsilon}_{it}-\varepsilon_{it}| &\leq \max_{i,t}\left|(\hat{\lambda}_i - \hat{H}^{-1}\lambda_i^0)'\hat{F}_t\right| + \max_{i,t}\left|\lambda_i^{0\prime}\{(\hat{H}')^{-1}\hat{F}_t - F_t^0\}\right| \\
&\leq \max_i\left\|\hat{\lambda}_i - \hat{H}^{-1}\lambda_i^0\right\|\max_t\left\|\hat{F}_t\right\| + \max_i\left\|\lambda_i^0\right\|\max_t\left\|\hat{F}_t - \hat{H}'F_t^0\right\|\left\|(\hat{H}')^{-1}\right\| \\
&= O_P(T^{-1/2}(\ln T)^{1/2})O_P(T^{\gamma_1/4}) + O_P(N^{\gamma_2/4})O_P((N/\ln T)^{1/2}) \\
&= O_P\left((T^{-1/2+\gamma_1/4} + N^{-1/2+\gamma_2/4})(\ln T)^{1/2}\right) = o_P(1),
\end{aligned}$$

where we use the fact that $\max_t\left\|\hat{F}_t\right\| \leq \max_t\left\|\hat{F}_t - \hat{H}'F_t^{(0)}\right\| + \max_t\left\|\hat{H}'F_t^{(0)}\right\| = O_P(T^{\gamma_1/4})$.

(iii) This follows from (i) and (ii) and Theorem 5 in Fan, Liao, and Mincheva (2013). ■

Proof of Lemma B.1. (i) Following the proof of Theorem 1 in Bai and Ng (2002) and that of Theorem 2.1, we can readily show that

$$\frac{1}{T}\left\|\breve{F}^R - F^0\breve{H}_{1R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right). \tag{C.5}$$

Recall that $\tilde{D}_R$ denotes the $R\times R$ diagonal matrix that contains the R largest eigenvalues of $(NTp^2)^{-1}X^*X^{*\prime}$ arranged in descending order along its diagonal line. Then $(NTp^2)^{-1}X^*X^{*\prime}\tilde{U}_R = \tilde{U}_R\tilde{D}_R$. This, along with the definition that $\breve{F}^R = \left(NTp^2\right)^{-1}X^*X^{*\prime}\tilde{F}^R$ and the fact that $\tilde{F}^R = \sqrt{T}\tilde{U}_R$, implies that

$$\breve{F}^R = \sqrt{T}\left(NTp^2\right)^{-1}X^*X^{*\prime}\tilde{U}_R = \sqrt{T}\tilde{U}_R\tilde{D}_R.$$

Then by (C.5), we have $\frac{1}{T}\left\|\sqrt{T}\tilde{U}_R\tilde{D}_R - F^0\breve{H}_{1R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$.

(ii) Following the proof of Theorem 1 in Bai and Ng (2002) and that of Theorem 2.1, we can readily show that $\frac{1}{N}\left\|\breve{\Lambda}^R - \Lambda^0\breve{H}_{2R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. Noting that $(NTp^2)^{-1}X^{*\prime}X^*\tilde{V}_R = \tilde{V}_R\tilde{D}_R$ and $\bar{\Lambda}^R = \sqrt{N}\tilde{V}_R$, we have

$$\breve{\Lambda}^R = \left(NTp^2\right)^{-1}X^*X^{*\prime}\bar{\Lambda}^R = \sqrt{N}\left(NTp^2\right)^{-1}X^*X^{*\prime}\tilde{V}_R = \sqrt{N}\tilde{V}_R\tilde{D}_R.$$

It follows that $\frac{1}{N}\left\|\sqrt{N}\tilde{V}_R\tilde{D}_R - \Lambda^0\breve{H}_{2R}\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$. ■

Proof of Lemma B.2. (i) Note that $\breve{\sigma}_r^2 = (NT)^{-1}\tilde{\sigma}_r^2$ denotes the rth largest eigenvalue of $(NTp^2)^{-1}X^*X^{*\prime}$. In view of that $X^* = X \circ G^* = \left(F^0\Lambda^{0\prime} + \varepsilon\right) \circ G^*$, we have

$$\begin{aligned}
(NTp^2)^{-1}X^*X^{*\prime} &= \frac{1}{NTp^2}\left[\left(F^0\Lambda^{0\prime} + \varepsilon\right) \circ G^*\right]\left[\left(F^0\Lambda^{0\prime} + \varepsilon\right) \circ G^*\right]' \\
&= \frac{1}{NTp^2}\left[\left(F^0\Lambda^{0\prime}\right) \circ G^*\right]\left[\left(F^0\Lambda^{0\prime}\right) \circ G^*\right]' + \frac{1}{NT}(\varepsilon \circ G^*)(\varepsilon \circ G^*)' \\
&\quad + \frac{1}{NT}\left[\left(F^0\Lambda^{0\prime}\right) \circ G^*\right](\varepsilon \circ G^*)' + \frac{1}{NT}(\varepsilon \circ G^*)\left[\left(F^0\Lambda^{0\prime}\right) \circ G^*\right]' \\
&\equiv IV_1 + IV_2 + IV_3 + IV_4.
\end{aligned}$$

As in the proof of Lemma A.1 and using Lemma C.4 below, it is easy to show that $\|\varepsilon \circ G^*\|_{\text{sp}} \leq p\,\|\varepsilon\|_{\text{sp}} + \|\varepsilon \circ [G^* - p\mathbf{1}_{T\times N}]\|_{\text{sp}} = O_P(\sqrt{N} + \sqrt{T})$. Then

$$\begin{aligned}
\|IV_2\|_{\text{sp}} &\leq \frac{1}{NTp^2}\|\varepsilon \circ G^*\|_{\text{sp}}^2 = O_P\left(\delta_{NT}^{-2}\right), \text{ and} \\
\|IV_3\|_{\text{sp}} &= \|IV_4\|_{\text{sp}} \leq \frac{1}{p^2\sqrt{NT}}\left\|F^0\Lambda^{0\prime}\right\|\frac{1}{\sqrt{NT}}\|\varepsilon \circ G^*\|_{\text{sp}} = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

For IV_1, we use $G^* = p\mathbf{1}_{T\times N} + (G^* - p\mathbf{1}_{T\times N})$ and make the following decomposition,

$$\begin{aligned}
IV_1 &= \frac{1}{NT}F^0\Lambda^{0\prime}\Lambda^0F^{0\prime} + \frac{1}{NTp^2}\left[\left(F^0\Lambda^{0\prime}\right) \circ (G^* - p\mathbf{1}_{T\times N})\right]\left[\left(F^0\Lambda^{0\prime}\right) \circ (G^* - p\mathbf{1}_{T\times N})\right]' \\
&\quad + \frac{1}{NTp}\left(F^0\Lambda^{0\prime}\right)\left[\left(F^0\Lambda^{0\prime}\right) \circ (G^* - p\mathbf{1}_{T\times N})\right]' + \frac{1}{NTp}\left[\left(F^0\Lambda^{0\prime}\right) \circ (G^* - p\mathbf{1}_{T\times N})\right]\Lambda^0F^{0\prime} \\
&\equiv IV_{1,1} + IV_{1,2} + IV_{1,3} + IV_{1,4}.
\end{aligned}$$

Using Lemma C.4 and following the analysis of $\left\|\left(F^0\Lambda^{0\prime}\right) \circ G\right\|$ in the proof of Lemma A.1, it is easy to show that $\left\|\left(F^0\Lambda^{0\prime}\right) \circ (G^* - p\mathbf{1}_{T\times N})\right\|_{\text{sp}} = O_P(\sqrt{N} + \sqrt{T})$, with which we can show that

$$\|IV_{1,2}\| = O_P\left(\delta_{NT}^{-2}\right) \text{ and } \|IV_{1,3}\| = \|IV_{1,4}\| = O_P\left(\delta_{NT}^{-1}\right).$$

Then by the Weyl's and triangular inequalities, we have

$$\left|\breve{\sigma}_r^2 - \mu_r\left(\frac{1}{NT}F^0\Lambda^{0\prime}\Lambda^0F^{0\prime}\right)\right| \leq \|IV_2 + IV_3 + IV_4 + IV_{1,2} + IV_{1,3} + IV_{1,4}\|_{\text{sp}} = O_P\left(\delta_{NT}^{-1}\right).$$

In addition, $\mu_r\left(\frac{1}{NT}F^0\Lambda^{0\prime}\Lambda^0F^{0\prime}\right) - \sigma_r^2 = O_P\left(\delta_{NT}^{-1}\right)$ under Assumption A.1(v). It follows that $\left|\breve{\sigma}_r^2 - \sigma_r^2\right| = O_P\left(\delta_{NT}^{-1}\right)$.

(ii) Let $\varepsilon^* = \frac{1}{p}\varepsilon \circ G^*$, $C^* = \frac{1}{p}(F^0\Lambda^{0\prime}) \circ [G^* - p\mathbf{1}_{T\times N}]$ and $\varsigma^* = C^* + \varepsilon^*$. Then

$$\frac{1}{p}X^* = \frac{1}{p}X \circ G^* = \frac{1}{p}\left(F^0\Lambda^{0\prime} + \varepsilon\right) \circ G^* = F^0\Lambda^{0\prime} + \varsigma^*.$$

Let $P_{\Lambda^0} = \Lambda^0(\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}$ and $Q_{\Lambda^0} = I_N - P_{\Lambda^0}$. Let $F^* = F^0 + \varsigma^*\Lambda^0(\Lambda^{0\prime}\Lambda^0)^{-1}$. Then

$$\frac{1}{NTp^2}X^*X^{*\prime} = \frac{1}{NT}F^{*\prime}\Lambda^{0\prime}\Lambda^0F^* + \frac{1}{NT}\varsigma^*Q_{\Lambda^0}\varsigma^{*\prime}.$$

It follows that for any $r \geq 1$

$$\begin{aligned} \breve{\sigma}^2_{R_0+r} &= \mu_{R_0+r}\left(\frac{1}{NTp^2}X^*X^{*\prime}\right) \leq \mu_{R_0+1}\left(\frac{1}{NT}F^{*\prime}\Lambda^{0\prime}\Lambda^0F^*\right) + \mu_r\left(\frac{1}{NT}\varsigma^*Q_{\Lambda^0}\varsigma^{*\prime}\right) \\ &= \mu_r\left(\frac{1}{NT}\varsigma^*Q_{\Lambda^0}\varsigma^{*\prime}\right), \end{aligned}$$

where we use the fact that $\text{rank}(F^{*\prime}\Lambda^{0\prime}\Lambda^0F^*) \leq R_0$. Using Lemma C.4, we can readily show that $\|\varsigma^*\|_{\text{sp}} = O_P(\sqrt{N}+\sqrt{T})$. Then

$$\mu_r\left(\frac{1}{NT}\varsigma^*Q_{\Lambda^0}\varsigma^{*\prime}\right) \leq \mu_r\left(\frac{1}{NT}\varsigma^*\varsigma^{*\prime}\right) \leq \frac{1}{NT}\|\varsigma^*\|^2_{\text{sp}} = O_P\left(\delta^{-2}_{NT}\right).$$

It follows that $\breve{\sigma}^2_{R_0+r} = O_P\left(\delta^{-2}_{NT}\right)$ for any $r \geq 1$.

(iii) To determine the lower probability bound for $\breve{\sigma}^2_{R_0+r}$, we notice that

$$\begin{aligned} \mu_{2R_0+r}\left(\frac{1}{NT}\varsigma^*\varsigma^{*\prime}\right) &\leq \mu_{R_0+r}\left(\frac{1}{NT}\varsigma^*Q_{\Lambda^0}\varsigma^{*\prime}\right) + \mu_{R_0+1}\left(\frac{1}{NT}\varsigma^*P_{\Lambda^0}\varsigma^{*\prime}\right) \\ &= \mu_{R_0+r}\left(\frac{1}{NT}\varsigma^*Q_{\Lambda^0}\varsigma^{*\prime}\right) \leq \mu_{R_0+r}\left(\frac{1}{NTp^2}X^*X^{*\prime}\right) = \breve{\sigma}^2_{R_0+r}. \end{aligned}$$

Without loss of generality we assume that $T \leq N$ and consider two cases: (1) T and N pass to infinity at the same rate (viz., $T \asymp N$) , and (2) $T = o(N)$. In Case (1), we can follow the proof of Lemma A.9 in Ahn and Horenstein (2013) to show that $\delta^2_{NT}\mu_{2R_0+r}\left(\frac{1}{NT}\varsigma^*\varsigma^{*\prime}\right)$ is bounded from below by a positive constant. In Case (2), we can consider the principal submatrix of ς^* and show that $\delta^2_{NT}\mu_{2R_0+r}\left(\frac{1}{NT}\varsigma^*\varsigma^{*\prime}\right)$ is also bounded from below by a positive constant. It follows that $\delta^2_{NT}\breve{\sigma}^2_{R_0+r}$ is bounded in probability from below by a positive constant, say c_σ, as $(N,T) \to \infty$. ■

Proof of Lemma B.3. Let $r \geq R_0+1$. Recall from the proof of Theorem 3.1 that $\tilde{F} = \tilde{F}^{R_0}$ and $\tilde{H} = \tilde{H}_{R_0}$. Note that

$$\left\|\frac{\tilde{u}_r'F^0}{\sqrt{T}}\right\| = \left\|\frac{\tilde{u}_r'F^0\tilde{H}}{\sqrt{T}}\tilde{H}^{-1}\right\| = \left\|\frac{\tilde{u}_r'(F^0\tilde{H}-\tilde{F})}{\sqrt{T}}\tilde{H}^{-1}\right\| \leq \left\|\tilde{H}^{-1}\right\|\|\tilde{u}_r\|\left\|\frac{F^0\tilde{H}-\tilde{F}}{\sqrt{T}}\right\| = O_P(\delta^{-1}_{NT}),$$

where the second inequality is by orthogonality between $\tilde{u}_r$ and $\tilde{F} = \tilde{F}^{R_0}$ for $r > R_0$. Analogously, we can show that $\frac{\tilde{v}_r'\Lambda^0}{\sqrt{N}} = O_P(\delta^{-1}_{NT})$. In the following, we aim at improving the probability order to show that $\tilde{v}_r'\Lambda^0 = O_P(\delta^{-1}_{NT})$ and $\tilde{u}_r'F^0 = O_P(\delta^{-1}_{NT})$

By the definition of singular value decomposition (SVD), we can write $\frac{1}{p}X^* = \sum_{k=1}^{N\wedge T}\tilde{u}_k\tilde{v}_k'\tilde{\sigma}_k$. Recall that $\varsigma^* \equiv \varepsilon \circ G^* + F^0\Lambda^{0\prime} \circ [G^* - E(G^*)]/p$, $\frac{1}{p}X^* = F^0\Lambda^{0\prime} + \varsigma^*$, and $\tilde{u}_r$ denotes the rth eigenvector of $\frac{1}{p^2}X^*X^{*\prime}$ that is associated with its rth largest eigenvalue. If follows that

$$\left(\frac{F^0\Lambda^{0\prime}\Lambda^0F^{0\prime}}{NT} + \frac{F^0\Lambda^{0\prime}\varsigma^{*\prime}}{NT} + \frac{\varsigma^*\Lambda^0F^{0\prime}}{NT} + \frac{\varsigma^*\varsigma^{*\prime}}{NT}\right)\tilde{u}_r = \tilde{u}_r\frac{\tilde{\sigma}^2_r}{NT}.$$

Premultiplying both sides of the above equation by $F^{0\prime}/\sqrt{T}$, we have

$$\frac{F^{0\prime}F^0}{T}\frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^{0\prime}\tilde{u}_r}{\sqrt{T}}+\frac{F^{0\prime}F^0}{T}\frac{\Lambda^{0\prime}\varsigma^{*\prime}\tilde{u}_r}{N\sqrt{T}}=O_P(\delta_{NT}^{-2}),$$

where we used the fact that $\frac{F^{0\prime}e\Lambda^0}{\sqrt{NT}}=O_P(1)$, $\frac{\tilde{\sigma}_r^2}{NT}=\breve{\sigma}_r^2=O_P(\delta_{NT}^{-2})$ for $r>R_0$, $\|\tilde{u}_r\|=1$, $\frac{1}{\sqrt{T}}\left\|F^{0\prime}\tilde{u}_r\right\|=O_P(\delta_{NT}^{-1})$ and $\left\|\varsigma^*/\sqrt{NT}\right\|_{\text{sp}}=O_P(\delta_{NT}^{-1})$. Premultiplying both sides of the above equation by $\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}$, we have

$$O_P(\delta_{NT}^{-2})=\frac{\Lambda^{0\prime}}{\sqrt{N}}\frac{\Lambda^0F^{0\prime}+\varsigma^{*\prime}}{\sqrt{NT}}\tilde{u}_r=\frac{\Lambda^{0\prime}}{\sqrt{N}}\frac{\frac{1}{p}X^{*\prime}\tilde{u}_r}{\sqrt{NT}}=\frac{\tilde{\sigma}_r}{\sqrt{NT}}\Lambda^{0\prime}\tilde{v}_r,$$

where the second equality follows from the decomposition $\frac{1}{p}X^*=F^0\Lambda^{0\prime}+\varsigma^*$ and the third one holds by the fact that $\frac{1}{p}X^{*\prime}\tilde{u}_r=\tilde{\sigma}_r\tilde{v}_r$. It follows that $\Lambda^{0\prime}\tilde{v}_r=O_P(\delta_{NT}^{-1})$ as $\frac{\delta_{NT}^{-}\tilde{\sigma}_r}{\sqrt{NT}}=\delta_{NT}^{-}\breve{\sigma}_r$ is bounded away from zero by Lemma B.2(iii). A symmetric argument gives that $\tilde{u}_r^{\prime}F^0=O_P(\delta_{NT}^{-1})$. ■

Proof of Theorem B.4. The proof follows closely from that of Theorem 1 in Negahban and Wainwright (2012). It suffices to show the probability of the event

$$\mathbf{E}_{NT}\equiv\left\{\exists\,\Gamma\in\mathcal{C}_{NT}(c_0)\mid\left|\left\|\frac{1}{\sqrt{p}}\Gamma\circ G\right\|-\|\Gamma\|\right|>\frac{7}{8}\|\Gamma\|+\frac{c_3\|\Gamma\|_\infty}{8}\right\}$$

is bounded by $c_1\exp(-c_2d\log d)$. Note that the claimed result holds for $c\Gamma$ too if it holds for Γ. In addition, since $\mathcal{C}_{NT}(c_0)$ is invariant to the rescaling of Γ, without loss of generality, we can prove the result by assuming that $\|\Gamma\|_\infty=\frac{1}{d}$. For any $\Gamma\in\mathcal{C}_{NT}(c_0)$ with $\|\Gamma\|_\infty=\frac{1}{d}$ and $\|\Gamma\|\le D$, we have $\|\Gamma\|_*\le\rho(D)$, where $\rho(D)\equiv\frac{D^2\sqrt{d}}{c_0\sqrt{\log d}}$ by the definition of $\mathcal{C}_{NT}(c_0)$. For each radius $D>0$, consider the set

$$\mathcal{B}(D)\equiv\left\{\Gamma\in\mathcal{C}_{NT}(c_0)\mid\|\Gamma\|_\infty=\frac{1}{d},\|\Gamma\|\le D,\|\Gamma\|_*\le\rho(D)\right\},$$

and the associated event

$$\mathbf{E}_{NT,D}\equiv\left\{\exists\,\Gamma\in\mathcal{B}(D)\mid\left|\left\|\frac{1}{\sqrt{p}}\Gamma\circ G\right\|-\|\Gamma\|\right|\ge\frac{3}{4}D+\frac{c_3}{8d}\right\}.$$

Lemma C.1 below shows that it suffices to obtain the upper bound for the probability of the event $\mathbf{E}_{NT,D}$ for each fixed $D>0$. In the second step, we show the probability of $\mathbf{E}_{NT,D}$ is bounded by $c_1\exp(-c_2D^2NT)$ for some universal constants (c_1,c_2).

Now, define

$$Z_{NT}(D)\equiv\sup_{\Gamma\in\overline{\mathcal{B}}(D)}\left|\left\|\frac{1}{\sqrt{p}}\Gamma\circ G\right\|-\|\Gamma\|\right|,$$

where $\overline{\mathcal{B}}(D)\equiv\left\{\Gamma\in\mathcal{C}_{NT}(c_0)\mid\|\Gamma\|_\infty\le\frac{1}{d},\|\Gamma\|\le D,\|\Gamma\|_*\le\rho(D)\right\}$. It suffices to show that there are universal constants (c_1,c_2,c_3) such that

$$P\left[Z_{NT}(D)\ge\frac{3}{4}D+\frac{c_3}{8d}\right]\le c_1\exp(-c_2D^2NT)\text{ for each fixed }D>0.$$

In order to prove the above result, we begin with a discretization argument. Let $\Gamma^1, \ldots, \Gamma^{N(\delta)}$ be a δ-covering of $\overline{\mathcal{B}}(D)$ in Frobenius norm. By definition, for any $\Gamma \in \overline{\mathcal{B}}(D)$, there exists some $k \in [N(\delta)]$ such that $\left\|\Gamma - \Gamma^k\right\| \leq \delta$. Let $\Delta \equiv \Gamma - \Gamma^k$. Then by the repeated use of the triangle inequality,

$$\begin{aligned} \left\|\frac{1}{\sqrt{p}}\Gamma \circ G\right\| - \|\Gamma\| &= \left\|\frac{1}{\sqrt{p}}(\Gamma^k + \Delta) \circ G\right\| - \left\|\Gamma^k + \Delta\right\| \\ &\leq \left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\| + \left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| + \|\Delta\| \\ &\leq \left|\left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| + \left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| + \delta. \end{aligned}$$

A symmetric argument gives the lower bound and establishes that this inequality holds for the absolute value of the difference:

$$\left|\left\|\frac{1}{\sqrt{p}}\Gamma \circ G\right\| - \|\Gamma\|\right| \leq \left|\left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| + \left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| + \delta.$$

Because both Γ and Γ^k belong to $\overline{\mathcal{B}}(D)$, we have that $\|\Delta\|_* \leq 2\rho(D)$ and $\|\Delta\|_\infty \leq 2/d$. Consequently, we have

$$Z_{NT}(D) \leq \delta + \max_{k \in [N(\delta)]} \left|\left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| + \sup_{\Delta \in \mathcal{D}(D,\delta)} \left|\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\|\right|,$$

where $\mathcal{D}(D, \delta) \equiv \left\{\Gamma \in \mathcal{C}_{NT}(c_0) \mid \|\Gamma\|_\infty \leq \frac{2}{d},\ \|\Gamma\| \leq \delta,\ \|\Gamma\|_* \leq 2\rho(D)\right\}$. Then by Lemmas C.2-C.3 below with the choice of $\delta = D/8$, we have

$$Z_{NT}(D) \leq \frac{D}{8} + \left(\frac{D}{8} + \frac{24}{d\sqrt{p}}\right) + \frac{D}{2} = \frac{3D}{4} + \frac{c_3}{8d},$$

with probability larger than $1 - c_1 \exp(-c_2 D^2 NT)$ by choosing large enough c_3. ■

The proof of Theorem B.4 relies on the following three lemmas whose proofs are given at the end of this section.

Lemma C.1 *Suppose that there are universal constants* (c_1, c_2) *such that* $P(\mathbf{E}_{NT,D}) \leq c_1 \exp(-c_2 D^2 NT)$ *for each fixed* $D > 0$*. Then there is a universal constant* c_2' *such that* $P(\mathbf{E}_{NT}) \leq c_1 \frac{\exp(-c_2' NT \log d/d)}{1-\exp(-c_2' NT \log d/d)}$.

Lemma C.2 *As long as* $d \geq 10$*, we have* $\max_{k \in [N(D/8)]} \left|\left\|\frac{1}{\sqrt{p}}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| \leq \frac{D}{8} + \frac{24}{d\sqrt{p}}$ *with probability greater than* $1 - 4\exp(-cd^2 \cdot D^2)$ *for some constant* $c > 0$.

Lemma C.3 $\sup_{\Delta \in \mathcal{D}(D,\delta)} \left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| \leq \frac{D}{2}$ *with probability at least* $1 - 2\exp(-\frac{pd^2D^2}{512})$.

To prove Theorem B.5, we need the following lemma.

Lemma C.4 *Let* $Z = \{Z_{it}\}$ *be a* $T \times N$ *matrix such that* Z_{it} *are independent across* (i,t)*,* $E(Z_{it}) = 0$*, and* $\max_{i,t} |Z_{it}| \leq c_c < \infty$ *with probability 1. Then there exists constants* M_1 *and* M_2 *such that for any* $t \geq 0$*,* $P\left(\|Z\|_{sp} \geq M_2(c_a \vee c_b) + t\right) \leq (N \wedge T)\exp\left(\frac{-t^2}{M_1 c_c^2}\right)$*, where* $c_a = \max_i \sqrt{\sum_{t=1}^T E\left(Z_{it}^2\right)}$ *and* $c_b = \max_t \sqrt{\sum_{i=1}^N E\left(Z_{it}^2\right)}$.

Proof. See Proposition 13 of Klopp (2015). ■

Proof of Theorem B.5. On the set $\mathcal{C}_{1NT}$, we define the metric $d(\cdot,\cdot)$ by the Frobenius norm, i.e., $d(\Gamma_1,\Gamma_2) \equiv \|\Gamma_1 - \Gamma_2\|$. For $\Gamma_1 = U_1V_1'$, $\Gamma_2 = U_2V_2' \in \mathcal{C}_{1NT}$, we have

$$\begin{aligned}
\|\Gamma_1 - \Gamma_2\|^2 &= \sum_{i=1}^{N}\sum_{t=1}^{T}(U_{1t}V_{1i} - U_{2t}V_{2i})^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}[(U_{1t} - U_{2t})V_{1i} + U_{2t}(V_{1i} - V_{2i})]^2 \\
&\leq 2\sum_{i=1}^{N} V_{1i}^2 \sum_{t=1}^{T}(U_{1t} - U_{2t})^2 + 2\sum_{i=1}^{N}(V_{1i} - V_{2i})^2 \sum_{t=1}^{T} {U_{2t}}^2 \\
&= 2(\|U_1 - U_2\|^2 + \|V_1 - V_2\|^2),
\end{aligned}$$

where the inequality holds by the fact $(a+b)^2 \leq 2(a^2+b^2)$ and the last equality is due to the fact $\|U_2\| = \|V_1\| = 1$. Let $\{U_l\}$ and $\{V_m\}$ be the minimum $\varepsilon/2$-nets of unit sphere in $\mathbb{R}^T$ and $\mathbb{R}^N$, respectively. Then for all $\Gamma = UV'$, there exists a pair (l,m) such that

$$\left\|\Gamma - U_lV_m'\right\|^2 \leq 2(\|U - U_l\|^2 + \|V - V_m\|^2) \leq \varepsilon^2.$$

Hence, $\{U_l\} \times \{V_m\}$ is an ε-net of $\mathcal{C}_{1NT}$. The covering number $\mathcal{N}(\mathcal{C}_{1NT}, d, \varepsilon)$ can be bounded by $\mathcal{N}(B_2^N, \|\cdot\|, \varepsilon/2) \times \mathcal{N}(B_2^T, \|\cdot\|, \varepsilon/2)$, where B_2^N denotes the unit ball in R^N space. By Corollary 4.2.13 of Vershynin (2018), we have $\mathcal{N}(\mathcal{C}_{1NT}, d, \varepsilon) \leq (6/\varepsilon)^{N+T}$. Let $\varepsilon_{NT} = 1/\log(N+T)$ and fix the minimum ε_{NT}-net $\{\Gamma_1, ..., \Gamma_K\}$ where $K \leq (6/\varepsilon_{NT})^{N+T}$. We have

$$\begin{aligned}
\sup_{\Gamma\in\mathcal{C}_{1NT}} \|\Gamma \circ [G - E(G)]\|_{\text{sp}} &\leq \max_{k\in\{1,...,K\}} \sup_{d(\Gamma,\Gamma_k)\leq\varepsilon_{NT}} \|\Gamma \circ [G - E(G)]\|_{\text{sp}} \\
&\leq \max_{k\in\{1,...,K\}} \left\{ \|\Gamma_k \circ [G - E(G)]\|_{\text{sp}} + \sup_{d(\Gamma,\Gamma_k)\leq\varepsilon_{NT}} \|(\Gamma - \Gamma_k) \circ [G - E(G)]\|_{\text{sp}} \right\} \\
&\leq \max_{k\in\{1,...,K\}} \|\Gamma_k \circ [G - E(G)]\|_{\text{sp}} + \max_{k\in\{1,...,K\}} \sup_{d(\Gamma,\Gamma_k)\leq\varepsilon_{NT}} \|\Gamma - \Gamma_k\| \\
&\leq \max_{k\in\{1,...,K\}} \|\Gamma_k \circ [G - E(G)]\|_{\text{sp}} + \varepsilon_{NT},
\end{aligned}$$

where the second inequality holds by the triangle inequality, the third inequality is due to the fact that $\|A\|_{\text{sp}} \leq \|A\|$ and every element of $G - E(G)$ is bounded by 1. For each k, we have $\Gamma_k = U_kV_k'$ for some unit vectors U_k and V_k. Let $Z^{(k)} \equiv \Gamma_k \circ [G - E(G)]$ and denote its (t,i)th entry as $Z_{it}^{(k)}$. By the definition of $\mathcal{C}_{1NT}$ and the fact that $G - E(G)$ has bounded i.i.d. entries, we can show

$$\begin{aligned}
\max\nolimits_{i,t}\left|Z_{it}^{(k)}\right| &\leq \|U_k\|_\infty \|V_k\|_\infty \leq c_{3NT}, \\
\max\nolimits_i \left(\sum_{t=1}^{T} E[(Z_{it}^{(k)})^2]\right)^{1/2} &\leq \|V_k\|_\infty \leq c_{2NT}, \text{ and } \max\nolimits_t \left(\sum_{i=1}^{N} E[(Z_{it}^{(k)})^2]\right)^{1/2} \leq \|U_k\|_\infty \leq c_{1NT}.
\end{aligned}$$

By Lemma C.4, there are some universal constants M_1 and M_2 such that

$$P\left(\left\|Z^{(k)}\right\|_{\text{sp}} \geq M_2(c_{1NT} \vee c_{2NT}) + t\right) \leq (N \wedge T)\exp\left(-\frac{t^2}{M_1c_{3NT}^2}\right).$$

Letting $t = KM_1^{1/2} c_{3NT}\sqrt{(N+T)\log\log(N+T)}$ and noting that $K = (6\log(N+T))^{N+T}$, we have

$$\begin{aligned} & P\left(\max_{k\in\{1,\ldots,K\}} \left\|Z^{(k)}\right\|_{\mathrm{sp}} \geq M_2(c_{1NT} \vee c_{2NT}) + t\right) \\ \leq\ & (6\log(N+T))^{N+T}\,(N\wedge T)\exp\left(-K^2(N+T)\log\log(N+T)\right) \\ =\ & \exp\left(-(K^2-1)(N+T)\log\log(N+T) + \log(N\wedge T) + (N+T)\log 6\right) \\ \leq\ & \exp\left(-(N+T)\log\log(N+T)\right), \end{aligned}$$

as long as $(K^2-3)\log\log(N+T) \geq \log 6$ and $\log(N+T) > (N\wedge T)^{1/(N+T)}$. Hence we have shown that

$$\max_{k\in\{1,\ldots,K\}} \left\|Z^{(k)}\right\|_{\mathrm{sp}} = O_P(c_{1NT} + c_{2NT} + c_{3NT}\sqrt{(N+T)\log\log(N+T)}).$$

To sum up, we have $\sup_{\Gamma\in\mathcal{C}_{1NT}} \|\Gamma\circ[G-E(G)]\| = O_P(c_{1NT}+c_{2NT}+c_{3NT}\sqrt{(N+T)\log\log(N+T)} + 1/\log(N+T))$. ■

Proof of Lemma C.1. For all $\Gamma \in \mathcal{C}_{NT}(c_0)$ with $\|\Gamma\|_\infty = \frac{1}{d}$, we have

$$\|\Gamma\|^2 \geq c_0\|\Gamma\|_* \sqrt{\frac{\log d}{d}} \geq c_0\|\Gamma\|\sqrt{\frac{\log d}{d}},$$

which implies that $\|\Gamma\| \geq \mu \equiv c_0\sqrt{\frac{\log d}{d}}$. Accordingly, recalling the definition (B.5), it suffices to restrict our attention to the sets $\mathcal{B}(D)$ with $D \geq \mu$. For $l = 1,2,\ldots$ and $\alpha = 7/6$, define the sets

$$\mathbb{S}_l \equiv \{\Gamma\in\mathcal{C}_{NT}(c_0) \mid \|\Gamma\|_\infty = \frac{1}{d},\ \|\Gamma\| \in [\alpha^{l-1}\mu, \alpha^l\mu],\ \text{and}\ \|\Gamma\|_* \leq \rho(\alpha^l\mu)\}.$$

Now, if the event $\mathbf{E}_{NT}$ holds for some matrix Γ, then $\Gamma\in\mathbb{S}_l \subset \overline{\mathcal{B}}(\alpha^l\mu)$ for some l and

$$\left|\left\|\frac{1}{\sqrt{p}}\Gamma\circ G\right\| - \|\Gamma\|\right| > \frac{7}{8}\|\Gamma\| + \frac{c_3\|\Gamma\|_\infty}{8} \geq \frac{7}{8}\alpha^{l-1}\mu + \frac{c_3\|\Gamma\|_\infty}{8} = \frac{3}{4}\alpha^l\mu + \frac{c_3}{8d},$$

where the equality holds by the fact that $\alpha = 7/6$ and $\|\Gamma\|_\infty = \frac{1}{d}$. Thus, $\mathbf{E}_{NT,\alpha^l\mu}$ occurs for some l. It follows that $\mathbf{E}_{NT} \subset \cup_{l=1}^\infty \mathbf{E}_{NT,\alpha^l\mu}$. By the union bound and the fact that $\alpha^{2l} \geq 2c^*l$ for some $c^* > 0$ and all $l \geq 1$, we have

$$\begin{aligned} P(\mathbf{E}_{NT}) &\leq \sum_{l=1}^\infty P(\mathbf{E}_{NT,\alpha^l\mu}) \leq c_1\sum_{l=1}^\infty \exp(-c_2\alpha^{2l}\mu^2 NT) \leq c_1\sum_{l=1}^\infty \exp(-2c^*c_2\mu^2 NTl) \\ &= c_1\sum_{l=1}^\infty \left[\exp(-2c^*c_2\mu^2 NT)\right]^l = c_1\frac{\exp(-c_2' NT\mu^2)}{1-\exp(-c_2' NT\mu^2)}, \end{aligned}$$

where the second inequality follows from the hypothesis on $P(\mathbf{E}_{NT,D})$ and $c_2' = 2c^*c_2$.

Since $NT\mu^2 = \frac{NT}{d}\log d$, the claim follows. ■

Proof of Lemma C.2. We first consider a fixed Γ and establish the exponential tail bound. Then we bound the covering number $N(D/8)$ and use the union bound to establish the result.

By the definition of Frobenius norm, we observe that for any $T \times N$ matrix A with typical element A_{it}, we have

$$\|A\| = \left[\sum_{i,t}(A_{it})^2\right]^{1/2} = \left[\sum_{i,t}(A_{it}z_{it})^2\right]^{1/2} = \sup_{\|U\|=1}\sum_{i,t}u_{it}A_{it}z_{it}$$

where z_{it}'s are i.i.d. Rademacher variables. Then

$$\left\|\frac{1}{\sqrt{p}}\Gamma \circ G\right\| = \left[\frac{1}{p}\sum_{i,t}(\Gamma_{it}g_{it})^2\right]^{1/2} = \sup_{\|U\|=1}\left(\sum_{i,t}u_{it}Y_{it}\right) \equiv Z_{NT},$$

where $Y_{it} \equiv \frac{1}{\sqrt{p}}z_{it}\Gamma_{it}g_{it}$, and z_{it}'s are i.i.d. Rademacher variables that are independent of $\{g_{it}\}$. Note that each Y_{it} is zero-mean, and bounded by $\frac{1}{\sqrt{p}d}$. By Corollary 4.8 in Ledoux (2001), we conclude that

$$P\left(|Z_{NT} - E(Z_{NT})| \geq \delta + \frac{8\sqrt{\pi}}{d\sqrt{p}}\right) \leq 4\exp(-\frac{p\delta^2 d^2}{8}), \text{ and } E(Z_{NT}^2) - [E(Z_{NT})]^2 \leq \frac{64}{pd^2}.$$

It follows that $\left|E(Z_{NT}) - \sqrt{E(Z_{NT}^2)}\right| \leq \frac{8}{\sqrt{p}d}$. With the above results and the fact that $E(Z_{NT}^2) = \|\Gamma\|^2$, we can conclude that

$$P\left(\left|\left\|\frac{1}{p}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| \geq \frac{D}{8} + \frac{24}{\sqrt{p}d}\right) \leq 4\exp(-\frac{pD^2d^2}{512}).$$

The upper bound of covering number $N(\delta)$ can be bounded similarly as in the proof of Lemma 4 in Negahban and Wainwright (2012). Then we have that

$$\log N(\delta) \leq 36(\rho(D)/\delta)^2 d, \text{ where } \rho(D) \equiv \frac{D^2\sqrt{d}}{c_0\sqrt{\log d}}.$$

Combining the tail bound with the union bound, we obtain

$$P\left(\max_{k\in[N(D/8)]}\left|\left\|\frac{1}{p}\Gamma^k \circ G\right\| - \left\|\Gamma^k\right\|\right| > \frac{D}{8} + \frac{24}{\sqrt{p}d}\right) \leq 4\exp(-\frac{pD^2d^2}{512} + 36(\rho(D)/\delta)^2 d).$$

Choosing the constant c_0 sufficiently large, we have the desired result. ■

Proof of Lemma C.3. Our goal is to bound the function $f(G) \equiv \sup_{\Delta\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\|$, where we recall that $\mathcal{D}(D,\delta) \equiv \left\{\Gamma \in \mathcal{C}_{NT}(c_0) | \left\|\Gamma\right\|_\infty \leq \frac{2}{d}, \|\Gamma\| \leq \delta, \|\Gamma\|_* \leq 2\rho(D)\right\}$.

(i) Our approach is to show concentration of G around its expectation $E[f(G)]$, and then upper bound the expectation. For any independent copy $\tilde{G}$ of G, we have

$$\begin{aligned}
f(G) - f(\tilde{G}) &= \sup_{\Delta\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| - \sup_{\tilde{\Delta}\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\tilde{\Delta} \circ \tilde{G}\right\| \\
&\leq \sup_{\Delta\in\mathcal{D}(D,\delta)}\left[\left\|\frac{1}{\sqrt{p}}\Delta \circ G\right\| - \left\|\frac{1}{\sqrt{p}}\Delta \circ \tilde{G}\right\|\right] \\
&\leq \sup_{\Delta\in\mathcal{D}(D,\delta)}\left\|\frac{1}{\sqrt{p}}\Delta \circ \left(G - \tilde{G}\right)\right\| \leq \frac{2}{\sqrt{p}d}\left\|G - \tilde{G}\right\|,
\end{aligned}$$

where the last inequality is by the fact $G-\tilde{G}$ has entries bounded by 1 and $\|\Delta\|_\infty \le \frac{2}{d}$. Therefore, by the bounded differences variant of the Azuma-Hoeffding inequality (Ledoux (2001, p.17)), we have $P\{|f(G)-E[f(G)]| \ge t\} \le 2\exp(-\frac{pd^2t^2}{8})$. Setting $t=\frac{D}{8}$, we have $P\{|f(G)-E[f(G)]| \ge \frac{D}{8}\} \le 2\exp(-\frac{pd^2D^2}{512})$.

(ii) Next we bound the expectation. First applying Jensen's inequality, we have

$$\begin{aligned}(E[f(G)])^2 &\le E[f^2(G)] = E\left(\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}\Delta_{it}^2\frac{g_{it}}{p}\right)\\ &= E\left\{\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}\left[\Delta_{it}^2\frac{g_{it}}{p}-E\left(\Delta_{it}^2\frac{g_{it}}{p}\right)\right]+\|\Delta\|^2\right\}\\ &\le E\left\{\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}\left[\Delta_{it}^2\frac{g_{it}}{p}-E\left(\Delta_{it}^2\frac{g_{it}}{p}\right)\right]\right\}+\delta^2,\end{aligned}$$

where we have used the fact that $\sum_{i,t}E\left(\Delta_{it}^2\frac{g_{it}}{p}\right)=\|\Delta\|^2\le\delta^2$. By a standard Rademacher symmetrization argument, we can show

$$E[f^2(G)] \le 2E\left[\sup_{\Delta\in\mathcal{D}(D,\delta)}\frac{1}{NT}\sum_{i,t}\left(NT\Delta_{it}^2\frac{g_{it}}{p}\xi_{it}\right)\right]+\delta^2,$$

where ξ_{it}s are i.i.d. Rademacher variables. Since $\left|NT\Delta_{it}^2\frac{g_{it}}{p}\xi_{it}\right| \le \frac{4NT}{d^2}$ for all (i,t), the Ledoux-Talagrand contraction inequality (e.g., Ledoux and Talagrand (1991, p.112)) implies that

$$E[f^2(G)] \le \frac{32\sqrt{NT}}{d^2\sqrt{p}}E\left[\sup_{\Delta\in\mathcal{D}(D,\delta)}\sum_{i,t}(\Delta_{it}g_{it}\xi_{it})\right]+\delta^2.$$

By the inequality that $|\mathrm{tr}(AB)| \le \|A\|_*\|B\|_{\mathrm{sp}}$, we have $\left|\sum_{i,t}(\Delta_{it}g_{it}\xi_{it})\right| \le \|\Delta\|_1\|G\circ\xi\|_{\mathrm{sp}}$. It follows that

$$E[f^2(G)] \le \frac{32\sqrt{NT}}{d^2\sqrt{p}}\rho(D)E\|G\circ\xi\|_{\mathrm{sp}}+\delta^2,$$

where we used the fact that $\|\Delta\|_* \le \rho(D)$. Noting that $G\circ\xi$ is a random matrix with bounded i.i.d. zero-mean entries, we have $E\|G\circ\xi\|_{\mathrm{sp}} \le \sqrt{d\log d}$; see, e.g. Theorem 4.4.5 of Vershynin (2018). Hence, we have

$$E[f(G)] \le \sqrt{E[f^2(G)]} \le \left(\frac{32\sqrt{NT}}{c_0d\sqrt{p}}D^2+\delta^2\right)^{1/2} \le \frac{7}{16}D,$$

by choosing a large enough c_0 and noting that $d=(N+T)/2\ge\sqrt{NT}$.

Combining the results of part (i)-(ii), we have the result desired. ■

D Additional Simulation Results

In this section, we report some additional simulation results.

D.1 Results for the 10% missing observations case ($q = 0.9$)

In this subsection, we report the simulation results for the case where $q = 0.9$, i.e., only 10% observations are missing at random. Tables A1–A2 correspond to Tables 2–3 in the main text. The results in Table A1 are comparable with those in Table 2. Our CV method performs slightly better than the case with a larger proportion of missing observations, and it continues to outperform the modified existing methods for most cases.

Table A1: Under/Over-estimation rate (%) with missing data ($q = 0.9$)

DGP	N	T	$\widetilde{CV}$	$\widehat{CV}$	M-ED	M-GR	M-ER	M-PC	M-IC
1	50	50	14.3/0.0	1.5/0.0	0.6/9.7	82.6/0.0	97.0/0.0	0.0/24.0	0.0/5.9
	50	100	0.6/0.0	0.1/0.0	0.0/4.5	71.6/0.0	93.9/0.0	0.0/6.7	0.0/3.0
	100	50	1.1/0.0	0.0/0.0	0.0/5.9	67.9/0.0	93.0/0.0	0.0/5.1	0.0/2.2
	100	100	0.0/0.0	0.0/0.0	0.0/1.4	37.6/0.0	83.3/0.0	0.0/2.3	0.0/1.0
2	50	50	9.4/0.0	0.1/0.1	0.6/4.9	79.6/0.0	95.4/0.0	0.0/86.1	0.0/38.0
	50	100	0.3/0.0	0.0/0.0	0.1/1.1	59.9/0.0	92.6/0.0	0.0/10.6	0.0/0.2
	100	50	0.5/0.0	0.0/1.0	0.0/0.8	65.5/0.0	93.0/0.0	0.0/84.4	0.0/25.2
	100	100	0.0/0.0	0.0/0.4	0.0/0.4	27.0/0.0	83.0/0.0	0.0/2.5	0.0/0.1
3	50	50	9.3/0.0	0.4/0.0	0.6/4.0	81.4/0.0	96.3/0.0	0.0/49.9	0.0/4.6
	50	100	0.9/0.0	0.0/0.0	0.0/0.8	63.4/0.0	92.8/0.0	0.0/9.8	0.0/0.0
	100	50	0.8/0.0	0.0/0.0	0.1/2.4	59.7/0.0	92.0/0.0	0.0/0.5	0.0/0.0
	100	100	0.0/0.0	0.0/0.0	0.0/0.4	28.3/0.0	81.1/0.0	0.0/0.0	0.0/0.0
4	50	50	10.0/0.0	0.2/0.0	0.9/3.9	76.0/0.0	95.0/0.0	0.0/27.6	0.0/1.1
	50	100	0.1/0.0	0.0/0.0	0.0/1.2	59.3/0.0	92.0/0.0	0.0/0.3	0.0/0.0
	100	50	1.0/0.0	0.0/0.0	0.0/2.6	61.2/0.0	90.1/0.0	0.0/0.4	0.0/0.0
	100	100	0.0/0.0	0.0/0.0	0.0/0.1	25.7/0.0	78.4/0.0	0.0/0.0	0.0/0.0
5	50	50	15.5/3.0	3.9/2.2	1.8/27.0	84.6/1.2	96.8/0.4	0.0/72.7	0.0/49.4
	50	100	0.9/3.8	0.2/2.6	0.0/19.6	76.0/0.9	94.5/0.5	0.0/56.2	0.0/42.3
	100	50	1.8/4.0	0.3/2.8	0.1/21.8	72.4/1.7	94.0/0.4	0.0/56.0	0.0/42.6
	100	100	0.1/2.4	0.0/1.4	0.0/17.8	48.0/1.6	86.8/0.6	0.0/47.7	0.0/37.9

Note: We report the under/over-estimation rate with missing data, where each entry is observed with probability $q = 0.9$. We consider $\widetilde{CV}$ and $\widehat{CV}$ with leave-out probability 1 - $p = 0.1$. For comparison, we also consider the M-ED,M-ER, M-PC, and M-IC, which are modified from ED of Onatski (2010), GR and ER of Ahn and Horenstein (2013), and PC and IC of Bai and Ng (2002), respectively. The number of replications is 1000.

The results in Table A2 are comparable with those in Table 3. As expected, the MSEs decrease as either N or T increases, and the MSEs in the case of $q = 0.9$ are smaller than those for $q = 0.7$.

D.2 Estimation and inference results with a known number of factors

In this subsection, we report the estimation and inference results of the 'infeasible' estimators, which are obtained by using the correct number of factors. Tables A3 and A4 reports the results for $q = 0.7$ and 0.9, respectively. Comparing the results in Tables A3 and A4 with those in Tables 3 and A2, we find that they are similar. A noticeable difference is the estimators of the common component based

Table A2: Mean squared error and coverage probability of confidence intervals with missing data ($q = 0.9$)

			MSE of $\hat{C}_{it}$			CP of Standard CI			CP of Robust CI		
DGP	N	T	$\hat{C}_{it}^{(oracle)}$	$\hat{C}_{it}^{(0)}$	$\hat{C}_{it}^{(\ell^*)}$	$\hat{F}_{t}^{(oracle)}$	$\hat{F}_{t}^{(0)}$	$\hat{F}_{t}^{(\ell^*)}$	$\hat{F}_{t}^{(oracle)}$	$\hat{F}_{t}^{(0)}$	$\hat{F}_{t}^{(\ell^*)}$
1	50	50	0.250	0.489	0.288	91.1%	93.5%	90.4%	93.4%	96.3%	93.2%
	50	100	0.186	0.355	0.210	91.3%	93.3%	90.9%	93.3%	95.6%	93.5%
	100	50	0.187	0.360	0.211	91.6%	94.4%	92.2%	92.8%	95.8%	92.9%
	100	100	0.124	0.234	0.139	94.0%	95.4%	92.7%	95.2%	96.7%	94.5%
2	50	50	0.284	0.502	0.311	87.6%	91.4%	86.7%	90.6%	94.1%	89.8%
	50	100	0.197	0.355	0.217	89.4%	93.4%	90.3%	92.3%	95.5%	93.0%
	100	50	0.229	0.391	0.250	88.3%	93.3%	89.6%	88.6%	95.0%	90.9%
	100	100	0.142	0.248	0.155	92.3%	95.0%	92.2%	92.9%	95.5%	93.1%
3	50	50	0.248	0.467	0.278	82.5%	89.3%	82.2%	89.1%	93.1%	87.7%
	50	100	0.192	0.348	0.212	81.8%	89.6%	82.6%	91.1%	93.8%	90.0%
	100	50	0.176	0.339	0.196	80.2%	86.7%	80.1%	86.5%	90.5%	85.4%
	100	100	0.121	0.227	0.134	83.5%	89.3%	84.2%	90.7%	93.1%	90.7%
4	50	50	0.255	0.473	0.284	82.3%	90.0%	82.4%	86.5%	92.7%	87.3%
	50	100	0.187	0.344	0.207	84.1%	90.6%	84.9%	87.9%	94.0%	88.7%
	100	50	0.193	0.355	0.213	86.4%	90.8%	87.0%	89.0%	92.5%	89.5%
	100	100	0.127	0.231	0.139	86.4%	91.5%	86.8%	89.8%	92.9%	89.6%
5	50	50	0.319	0.568	0.359	91.9%	94.2%	91.3%	93.0%	96.0%	93.1%
	50	100	0.255	0.421	0.265	92.1%	93.5%	91.1%	93.3%	95.5%	93.9%
	100	50	0.258	0.428	0.267	92.4%	94.3%	92.4%	93.7%	95.8%	93.6%
	100	100	0.156	0.266	0.165	94.6%	95.5%	92.8%	95.3%	96.7%	94.6%

Note: We report the mean squared errors (MSE) of $\hat{C}_{it}$ and the coverage probabilities (CP) of the 95% confidence intervals (CIs) for F_t^0's. Each entry is observed with probability $q = 0.9$. We consider the feasible estimates with $\ell = 0$ and $\ell = \ell^*$ and the oracle estimate that is obtained using the information of R^0 and missing observations. The standard CI's and the robust CI's are constructed using $\hat{\Gamma}_{1g,t}^{(1)}$ and $\hat{\Gamma}_{1g,t}^{(2)}$ in Section 2.4, respectively.

on the true number of factors have slightly slower MSE than those based based on the estimated number of factors.

D.3 Determining the number of factors with heterogeneous missing

In this subsection, we evaluate the methods of determining the number of factors when the missing is heterogeneous. Specifically, we consider two missing mechanisms that are also considered by Zhu, Wang and Samworth (2019):

> **Case A**: $P(g_{it} = 1) = p_i q_t$ for all (i, t)'s, where p_i's are i.i.d. U[0.5, 1] and q_t's are i.i.d. U[0.5, 1];
>
> **Case B**: $P(g_{it} = 1) = 0.3$ for $i \le N/2$, and $g_{it} = 1$ for all $i > N/2$.

The DGPs are the same as those considered in Section 4.

Table A5 presents the under/over-estimation rate in case A. Thus, all entries of X are observed with a different probability. In this case, $\widehat{CV}$ continue to accurately estimate the number of factors. $\widetilde{CV}$ and M-IC show an obvious pattern of convergence. M-ED shows a pattern of convergence but has considerable under-estimation rate for all DGPs. Both M-GR and M-ER almost always under-estimate the number of factors. M-PC tends to over-estimate the number of factors.

Table A3: Mean squared errors and coverage probability of confidence intervals of infeasible estimates with missing data ($q = 0.7$)

			MSE of $\hat{C}_{it}$			CP of Standard CI			CP of Robust CI		
DGP	N	T	$\hat{C}_{it}^{(oracle)}$	$\hat{C}_{it}^{(0)}$	$\hat{C}_{it}^{(\ell^*)}$	$\hat{F}_t^{(oracle)}$	$\hat{F}_t^{(0)}$	$\hat{F}_t^{(\ell^*)}$	$\hat{F}_t^{(oracle)}$	$\hat{F}_t^{(0)}$	$\hat{F}_t^{(\ell^*)}$
1	50	50	0.250	1.493	0.425	91.1%	96.2%	86.0%	93.4%	98.4%	90.4%
	50	100	0.186	0.994	0.290	91.3%	94.3%	89.6%	93.3%	96.4%	92.5%
	100	50	0.187	1.042	0.293	91.6%	95.9%	91.1%	92.8%	97.7%	92.1%
	100	100	0.124	0.613	0.185	94.0%	95.5%	91.8%	95.2%	96.6%	94.1%
2	50	50	0.284	1.406	0.414	87.6%	94.8%	84.1%	90.6%	97.8%	88.6%
	50	100	0.197	0.933	0.282	89.4%	95.7%	87.7%	92.3%	97.0%	91.0%
	100	50	0.229	0.997	0.312	88.3%	96.0%	87.8%	88.6%	98.1%	91.0%
	100	100	0.142	0.604	0.193	92.3%	96.9%	92.0%	92.9%	98.1%	93.4%
3	50	50	0.248	1.388	0.388	82.5%	93.4%	79.8%	89.1%	97.2%	85.4%
	50	100	0.192	0.928	0.278	81.8%	92.8%	82.3%	91.1%	96.3%	88.3%
	100	50	0.176	0.963	0.263	80.2%	94.3%	80.1%	86.5%	96.8%	84.4%
	100	100	0.121	0.588	0.173	83.5%	93.1%	86.4%	90.7%	96.1%	90.4%
4	50	50	0.255	1.391	0.393	82.3%	93.7%	81.5%	86.5%	98.0%	86.5%
	50	100	0.187	0.934	0.273	84.1%	94.4%	85.1%	87.9%	97.6%	89.2%
	100	50	0.193	0.977	0.280	86.4%	95.8%	87.9%	89.0%	97.4%	90.3%
	100	100	0.127	0.591	0.179	86.4%	93.4%	87.0%	89.8%	95.4%	90.1%
5	50	50	0.319	1.614	0.523	91.9%	96.9%	86.2%	93.0%	98.5%	90.5%
	50	100	0.255	1.126	0.382	92.1%	94.4%	90.2%	93.3%	96.2%	93.1%
	100	50	0.258	1.177	0.387	92.4%	96.2%	91.1%	93.7%	98.2%	92.0%
	100	100	0.156	0.695	0.238	94.6%	95.7%	92.3%	95.3%	96.9%	93.7%

Note: We report the mean squared errors (MSE) of $\hat{C}_{it}$ and the coverage probabilities (CP) of the 95% confidence intervals (CIs) for F_t^0's. Each entry is observed with probability $q = 0.7$. We consider the feasible estimates with $\ell = 0$ and $\ell = \ell^*$, and the oracle estimate that is obtained using the information of R^0 and missing observations. The standard CI's and the robust CI's are constructed using $\hat{\Gamma}_{1g,t}^{(1)}$ and $\hat{\Gamma}_{1g,t}^{(2)}$ in Section 2.4, respectively.

Table A4: Mean squared errors and coverage probability of confidence intervals of infeasible estimates with missing data ($q = 0.9$)

			MSE of $\hat{C}_{it}$			CP of Standard CI			CP of Robust CI		
DGP	N	T	$\hat{C}_{it}^{(oracle)}$	$\hat{C}_{it}^{(0)}$	$\hat{C}_{it}^{(\ell^*)}$	$\hat{F}_t^{(oracle)}$	$\hat{F}_t^{(0)}$	$\hat{F}_t^{(\ell^*)}$	$\hat{F}_t^{(oracle)}$	$\hat{F}_t^{(0)}$	$\hat{F}_t^{(\ell^*)}$
1	50	50	0.250	0.488	0.285	91.1%	93.5%	90.4%	93.4%	96.2%	93.2%
	50	100	0.186	0.356	0.210	91.3%	93.3%	90.9%	93.3%	95.6%	93.5%
	100	50	0.187	0.360	0.211	91.6%	94.4%	92.2%	92.8%	95.8%	92.9%
	100	100	0.124	0.234	0.139	94.0%	95.4%	92.7%	95.2%	96.7%	94.5%
2	50	50	0.284	0.502	0.311	87.6%	91.4%	86.7%	90.6%	94.1%	89.8%
	50	100	0.197	0.355	0.217	89.4%	93.4%	90.3%	92.3%	95.5%	93.0%
	100	50	0.229	0.389	0.248	88.3%	93.3%	89.6%	88.6%	95.0%	90.9%
	100	100	0.142	0.247	0.154	92.3%	95.0%	92.2%	92.9%	95.5%	93.1%
3	50	50	0.248	0.467	0.277	82.5%	89.3%	82.2%	89.1%	93.1%	87.7%
	50	100	0.192	0.348	0.212	81.8%	89.6%	82.6%	91.1%	93.8%	90.0%
	100	50	0.176	0.339	0.196	80.2%	86.7%	80.1%	86.5%	90.5%	85.4%
	100	100	0.121	0.227	0.134	83.5%	89.3%	84.2%	90.7%	93.1%	90.7%
4	50	50	0.255	0.473	0.284	82.3%	90.0%	82.4%	86.5%	92.7%	87.2%
	50	100	0.187	0.344	0.207	84.1%	90.6%	84.9%	87.9%	94.0%	88.7%
	100	50	0.193	0.355	0.213	86.4%	90.8%	87.0%	89.0%	92.5%	89.5%
	100	100	0.127	0.231	0.139	86.4%	91.5%	86.8%	89.8%	92.9%	89.6%
5	50	50	0.319	0.580	0.366	91.9%	94.0%	91.0%	93.0%	96.0%	93.1%
	50	100	0.255	0.429	0.274	92.1%	93.5%	91.1%	93.3%	95.5%	94.0%
	100	50	0.258	0.440	0.279	92.4%	94.4%	92.4%	93.7%	95.9%	93.7%
	100	100	0.156	0.273	0.173	94.6%	95.5%	92.9%	95.3%	96.7%	94.6%

Note: We report the mean squared errors (MSE) of $\hat{C}_{it}$ and the coverage probabilities (CP) of the 95% confidence intervals (CIs) for F_t^0's. Each entry is observed with probability $q = 0.9$. We consider the infeasible estimates with $\ell = 0$ and $\ell = \ell^*$ using correct number of factors. We consider the feasible estimates with $\ell = 0$ and $\ell = \ell^*$, and the oracle estimate that is obtained using the information of R^0 and missing observations. The standard CI's and the robust CI's are constructed using $\hat{\Gamma}_{1g,t}^{(1)}$ and $\hat{\Gamma}_{1g,t}^{(2)}$ in Section 2.4, respectively.

Table A5: Under/Over-estimation rate (%) with heterogeneous missing data (Case A)

DGP	N	T	$\widetilde{CV}$	$\widehat{CV}$	M-ED	M-GR	M-ER	M-PC	M-IC
1	50	50	81.3/0.2	28.5/0.0	77.4/6.2	99.6/0.0	99.9/0.0	0.0/94.8	0.0/77.9
	50	100	55.4/0.1	2.0/0.0	68.6/3.4	99.9/0.0	100.0/0.0	0.0/74.1	0.0/41.3
	100	50	59.4/0.5	3.4/0.0	69.0/5.2	99.4/0.0	100.0/0.0	0.0/68.6	0.0/35.5
	100	100	14.8/0.3	0.0/0.0	37.0/1.6	99.9/0.0	99.9/0.0	0.0/33.0	0.0/14.6
2	50	50	77.7/0.2	17.4/0.0	74.4/5.7	99.9/0.0	100.0/0.0	0.0/84.8	0.0/62.3
	50	100	46.5/0.0	0.5/0.0	63.1/2.7	99.9/0.0	100.0/0.0	0.0/42.8	0.0/15.4
	100	50	53.3/0.1	1.8/0.0	66.1/4.8	99.7/0.0	99.9/0.0	0.0/61.1	0.0/34.0
	100	100	10.3/0.2	0.0/0.0	29.4/1.0	100.0/0.0	100.0/0.0	0.0/18.5	0.0/5.6
3	50	50	80.6/0.0	20.4/0.0	78.6/5.7	99.9/0.0	100.0/0.0	0.0/83.6	0.0/57.4
	50	100	46.9/0.0	1.2/0.0	64.0/2.7	100.0/0.0	100.0/0.0	0.0/50.2	0.0/19.4
	100	50	55.0/0.3	2.1/0.0	67.0/4.0	99.7/0.0	99.8/0.0	0.0/30.8	0.0/9.6
	100	100	9.8/0.2	0.0/0.0	33.7/1.0	99.9/0.0	100.0/0.0	0.0/6.9	0.0/1.0
4	50	50	77.6/0.1	20.0/0.0	75.2/5.5	99.6/0.0	99.8/0.0	0.0/79.4	0.1/47.6
	50	100	48.7/0.2	1.2/0.0	66.5/3.0	99.7/0.0	100.0/0.0	0.0/32.6	0.0/7.7
	100	50	55.4/0.1	1.9/0.0	65.1/3.3	99.8/0.0	100.0/0.0	0.0/26.4	0.0/8.1
	100	100	10.7/0.0	0.0/0.0	31.0/1.2	99.8/0.0	100.0/0.0	0.0/3.1	0.0/0.5
5	50	50	79.5/1.5	33.2/1.3	74.2/9.2	99.6/0.0	99.8/0.0	0.0/97.0	0.0/90.6
	50	100	54.9/1.8	4.6/1.7	67.8/5.0	99.9/0.0	100.0/0.0	0.0/91.5	0.0/81.7
	100	50	58.6/2.0	6.1/2.1	68.0/7.2	99.7/0.0	100.0/0.0	0.0/89.9	0.0/77.5
	100	100	14.6/4.9	0.1/0.9	38.2/4.7	99.7/0.0	99.9/0.0	0.0/84.7	0.0/74.0

Note: We report the under/over-estimation rate with heterogeneous missing data in Case A. We consider $\widetilde{CV}$ and $\widehat{CV}$ with leave-out probability 1 - $p = 0.1$. For the comparison purpose, we also consider the M-ED,M-ER, M-PC, and M-IC, which are modified from ED of Onatski (2010), GR and ER of Ahn and Horenstein (2013), and PC and IC of Bai and Ng (2002), respectively. The number of replications is 1000.

Table A6 presents the under/over-estimation rate in case B. Thus, only half of the columns contain missing observations while the other half do not. The CV methods, M-ED, M-PC and M-IC perform well. $\widehat{CV}$ again dominates the other methods for all cases. M-GR and M-ER continue to severely under-estimate the number of factors.

In both cases, when the missing at random assumption fails, the CV methods still provide accurate estimation results. We conjecture that our theoretical analysis can be extended to allow more general missing mechanisms. It is an important and exciting topic, which is left for future research.

REFERENCES

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203-1227.

Allez, R., Bouchaud, J.-P., 2013. Eigenvector dynamics: general theory and some applications. Physical Review E, 86(4):046202.

Bai, J., 2003. Inferential theory for factors models of large dimensions. Econometrica 71, 135-173.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Bai, J. and Ng, S., 2019. Rank regularized estimation of approximate factor models. Journal of econometrics, 212, 78-96.

Table A6: Under/Over-estimation rate (%) with heterogeneous missing data (Case B)

DGP	N	T	$\widetilde{CV}$	$\widehat{CV}$	M-ED	M-GR	M-ER	M-PC	M-IC
1	50	50	16.2/2.4	16.3/0.6	13.9/22.8	96.6/0.0	99.3/0.0	0.0/45.2	0.1/20.6
	50	100	2.5/2.3	2.9/0.2	6.9/16.8	93.8/0.0	98.7/0.0	0.0/11.1	0.0/5.6
	100	50	2.1/0.8	1.1/0.0	3.1/11.0	91.2/0.0	97.6/0.0	0.0/12.9	0.0/6.2
	100	100	0.0/2.6	0.0/0.0	0.1/4.9	81.7/0.0	96.8/0.0	0.0/3.6	0.0/2.4
2	50	50	10.6/2.0	10.7/0.8	14.1/14.2	93.8/0.0	97.9/0.0	0.0/53.9	0.1/23.2
	50	100	1.5/1.9	1.3/0.3	5.2/12.3	91.7/0.0	98.9/0.0	0.0/4.3	0.0/0.6
	100	50	1.0/2.5	0.5/0.1	2.7/5.6	90.1/0.0	97.3/0.0	0.0/33.1	0.0/15.5
	100	100	0.0/7.8	0.0/0.0	0.1/2.3	78.6/0.0	97.4/0.0	0.0/1.0	0.0/0.2
3	50	50	10.5/1.1	11.8/0.9	13.5/15.2	95.4/0.0	99.0/0.0	0.0/58.0	0.0/27.9
	50	100	1.1/2.2	1.3/0.4	6.2/9.0	92.9/0.0	99.0/0.0	0.0/19.2	0.0/5.0
	100	50	1.3/0.6	0.6/0.0	2.4/6.3	90.3/0.0	97.5/0.0	0.0/4.6	0.0/0.9
	100	100	0.0/7.8	0.0/0.0	0.0/2.4	81.4/0.0	97.8/0.0	0.0/0.2	0.0/0.0
4	50	50	11.6/1.6	11.8/0.8	14.3/16.2	94.0/0.0	98.0/0.0	0.0/32.8	0.0/11.4
	50	100	0.6/1.6	0.7/0.1	5.5/14.1	92.4/0.0	99.0/0.0	0.0/1.8	0.0/0.3
	100	50	1.7/1.5	0.5/0.0	2.7/7.6	87.4/0.0	97.4/0.0	0.0/2.8	0.0/0.1
	100	100	0.0/5.6	0.0/0.0	0.1/2.5	79.8/0.0	96.3/0.0	0.0/0.0	0.0/0.0
5	50	50	16.5/8.2	18.7/2.6	15.8/29.6	96.6/0.4	99.3/0.1	0.0/71.7	0.0/54.9
	50	100	2.6/8.7	3.2/2.4	7.1/21.6	93.8/0.5	98.8/0.1	0.0/51.3	0.0/42.4
	100	50	2.8/7.4	1.8/2.4	3.6/20.7	91.3/0.7	97.8/0.2	0.0/56.5	0.0/44.1
	100	100	0.0/7.9	0.1/1.5	0.2/15.0	83.8/1.1	97.7/0.3	0.0/47.2	0.0/41.9

Note: We report the under/over-estimation rate with heterogeneous missing data in Case B. We consider $\widetilde{CV}$ and $\widehat{CV}$ with leave-out probability 1 - $p = 0.1$. For the comparison purpose, we also consider the M-ED,M-ER, M-PC, and M-IC, which are modified from ED of Onatski (2010), GR and ER of Ahn and Horenstein (2013), and PC and IC of Bai and Ng (2002), respectively. The number of replications is 1000.

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

Klopp, O., 2015. Matrix completion by singular value thresholding: sharp bounds. Electronic Journal of Statistics 9, 2348-2369.

Ledoux, M., 2001. The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs Number 89. American Mathematical Society.

Ledoux, M. and Talagrand, M., 2013. Probability in Banach Spaces: Isoperimetry and Processes. Springer Science & Business Media.

Negahban, S. and Wainwright, M.J., 2012. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. Journal of Machine Learning Research 13, 1665-1697.

Stewart, G. W., Sun, J.-G., 1990. Matrix Perturbation Theory. Academic Press, Boston.

Su, L., Wang, X., 2017. On time-varying factor models: estimation and testing. Journal of Econometrics 198, 84-101.

Vershynin, R., 2018. High-dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, New York.

Zhu, Z., Wang, T. and Samworth, R.J., 2019. High-dimensional principal component analysis with heterogeneous missingness. arXiv preprint arXiv:1906.12125.