Supplement to "Inference for Low-rank Completion without Sample Splitting with Application to Treatment Effect Estimation"

July 2023

This supplementary appendix provides some proofs that are relegated from the main text.

Contents

F Proofs of Main Results

F.1 β Estimation

Based on Proposition E.1, we can estimate the loadings unbiasedly.

Proposition F.1. *Suppose that Assumptions 3.2, 3.3, C.1 - C.3 hold. Then,*

$$\sqrt{T}\left(\widehat{\beta}_i - H_2'^{-1}\beta_i\right) = \sqrt{T}H_2'^{-1}\left(\sum_{s=1}^{T}\omega_{is}F_sF_s'\right)^{-1}\left(\sum_{s=1}^{T}\omega_{is}F_s\varepsilon_{is}\right) + \sqrt{T}R_i^{\beta},$$

$$\begin{aligned}\max_i \|\sqrt{T}R_i^{\beta}\| = O_P\Bigg(&\frac{\sigma^4 p_{\max}^2 c_{\text{inv}} q^{\frac{5}{2}} K^{\frac{1}{2}} \max\{N,T\} NT^{\frac{3}{2}}}{\psi_{\min}^4 p_{\min}^4} + \frac{\sigma^3 p_{\max}^{\frac{5}{2}} \vartheta c_{\text{inv}}^2 q^5 \mu K^2 \max\{N\sqrt{\log N}, T\sqrt{\log T}\}\sqrt{N}T}{\psi_{\min}^3 p_{\min}^4} \\ &+ \frac{\sigma^2 p_{\max}^{\frac{7}{2}} \vartheta c_{\text{inv}}^2 q^6 \mu^2 K^3 \max\{N\sqrt{\log N}, T\sqrt{\log T}\}\sqrt{T}}{\psi_{\min}^2 p_{\min}^5} + \frac{\sigma^2 \mu^{\frac{1}{2}} K^{\frac{1}{2}} \max\{N\log^2 N, T\log^2 T\}\sqrt{T}}{\psi_{\min}^2 p_{\min}^2} \\ &+ \frac{\sigma p_{\max}^{\frac{5}{2}} \vartheta c_{\text{inv}} q^{\frac{13}{2}} \mu^{\frac{5}{2}} K^{\frac{7}{2}} \max\{\sqrt{N\log N}, \sqrt{T\log T}\}\sqrt{T}}{\psi_{\min} p_{\min}^4 \min\{\sqrt{N}, \sqrt{T}\}} + \left[\frac{qK^{\frac{1}{2}}\sqrt{NT}}{\psi_{\min}} + \frac{\sigma p_{\max}^{\frac{1}{2}} NT^{\frac{3}{2}}}{\psi_{\min}^2 p_{\min}}\right]\max_{it}\left|M_{it}^R\right|\Bigg).\end{aligned}$$

Proof of Proposition F.1 As noted in the proof of Proposition E.1, we have the following decomposition:

$$\widehat{F}_s - H_2F_s = \widetilde{B}_s^{-1}H_1'\frac{1}{N}\left(\sum_{j=1}^{N}\omega_{js}\beta_j\varepsilon_{js}\right) + \sum_{d=1}^{4}\Delta_{d,s}, \tag{F.1}$$

$$\Delta_{1,s} := -B^{-1}H_1'\frac{1}{N}\sum_{j=1}^{N}(\omega_{js}-p_j)\beta_jF_s'H_1'^{-1}\left(\widetilde{\beta}_j - H_1'\beta_j\right) + \widetilde{B}_s^{-1}\frac{1}{N}\sum_{j=1}^{N}\omega_{js}\varepsilon_{js}\left(\widetilde{\beta}_j - H_1'\beta_j\right),$$

$$\Delta_{2,s} := \left(\widetilde{B}_s^{-1} - B^{-1}\right)\frac{1}{N}\sum_{j=1}^{N}\omega_{js}\widetilde{\beta}_j\left(\beta_j'H_1 - \widetilde{\beta}_j'\right)H_1^{-1}F_s,$$

$$\Delta_{3,s} := B^{-1}\frac{1}{N}\sum_{j=1}^{N}\omega_{js}\left(\widetilde{\beta}_j - H_1'\beta_j\right)\left(\beta_j'H_1 - \widetilde{\beta}_j'\right)H_1^{-1}F_s, \quad \Delta_{4,s} := \widetilde{B}_s^{-1}\frac{1}{N}\sum_{j=1}^{N}\omega_{js}\widetilde{\beta}_jM_{js}^R.$$

Define $\widehat{A}_i := \frac{1}{T}\sum_{s=1}^{T}\omega_{is}\widehat{F}_s\widehat{F}_s'$, $A_i := \frac{1}{T}\sum_{s=1}^{T}\omega_{is}H_2F_sF_s'H_2'$. By definition, $\widehat{\beta}_i = \left(\sum_{s=1}^{T}\omega_{is}\widehat{F}_s\widehat{F}_s'\right)^{-1}\sum_{s=1}^{T}\omega_{is}\widehat{F}_sy_{is}$. Substituting (F.1), basic algebras show the following identity:

$$\widehat{\beta}_i - H_2'^{-1}\beta_i = H_2'^{-1}\left(\sum_{s=1}^{T}\omega_{is}F_sF_s'\right)^{-1}\left(\sum_{s=1}^{T}\omega_{is}F_s\varepsilon_{is}\right) + \sum_{d=1}^{6}\delta_{d,i},$$

$$\delta_{1,i} := \widehat{A}_i^{-1}\frac{1}{T}\sum_{s=1}^{T}\omega_{is}\widetilde{B}_s^{-1}H_1'\frac{1}{N}\sum_{j=1}^{N}\omega_{js}\beta_j\varepsilon_{js}\varepsilon_{is}, \quad \delta_{2,i} := -\widehat{A}_i^{-1}\frac{1}{T}\sum_{s=1}^{T}\omega_{is}\widehat{F}_s\frac{1}{N}\sum_{j=1}^{N}\omega_{js}\varepsilon_{js}\beta_j'H_1\widetilde{B}_s^{-1}H_2'^{-1}\beta_i,$$

$$\delta_{3,i} := -\sum_{d=1}^{4}\widehat{A}_i^{-1}\frac{1}{T}\sum_{s=1}^{T}\omega_{is}\widehat{F}_s\Delta_{d,s}'H_2'^{-1}\beta_i, \quad \delta_{4,i} := \sum_{d=1}^{4}\widehat{A}_i^{-1}\frac{1}{T}\sum_{s=1}^{T}\omega_{is}\Delta_{d,s}\varepsilon_{is},$$

$$\delta_{5,i} := \left(\widehat{A}_i^{-1} - A_i^{-1}\right)H_2\frac{1}{T}\sum_{s=1}^{T}\omega_{is}F_s\varepsilon_{is}, \quad \delta_{6,i} := \widehat{A}_i^{-1}\frac{1}{T}\sum_{s=1}^{T}\omega_{is}\widehat{F}_sM_{is}^R.$$

Step 1. First, we bound $\max_i \|\delta_{1,i}\|$. Define $\gamma_s := \sum_{j=1}^{N}\omega_{js}\varepsilon_{js}(H_1'\beta_j)$. Then, we have

$$\begin{aligned}\delta_{1,i} = &\frac{1}{NT}\widehat{A}_i^{-1}\sum_{s=1}^{T}\left(\widetilde{B}_s^{-1} - B^{-1}\right)\gamma_s\omega_{is}\varepsilon_{is} + \frac{1}{NT}\widehat{A}_i^{-1}B^{-1}\sum_{s=1}^{T}\left(\sum_{j=1}^{N}(\varepsilon_{is}\varepsilon_{js} - \mathbb{E}[\varepsilon_{is}\varepsilon_{js}|\Omega,\mathcal{M}])\omega_{is}\omega_{js}H_1'\beta_j\right) \\ &+ \frac{1}{NT}\widehat{A}_i^{-1}B^{-1}\sum_{s=1}^{T}\left(\sum_{j=1}^{N}\mathbb{E}[\varepsilon_{is}\varepsilon_{js}|\Omega,\mathcal{M}]\omega_{is}\omega_{js}H_1'\beta_j\right).\end{aligned}$$

Because $\sum_{s=1}^T \|\gamma_s\|^2 = O_P\left(\sigma^2 p_{\max}\mu K \max\{N^2, T^2\}\right)$ by Claim F.6 (iv), the first term is bounded like

$$\frac{1}{NT}\max_i \|\widehat{A}_i^{-1}\| \max_s \|\widetilde{B}_s^{-1} - B^{-1}\| \left(\sum_{s=1}^T \|\gamma_s\|^2\right)^{\frac{1}{2}} \max_i \left(\sum_{s=1}^T \varepsilon_{is}^2\right)^{\frac{1}{2}}$$
$$= O_P\left(\frac{\sigma^3 p_{\max}^{\frac{5}{2}} c_{\text{inv}} q^2 \mu^{\frac{1}{2}} K \max\{N^{\frac{3}{2}}, T^{\frac{3}{2}}\}\sqrt{T}}{\psi_{\min}^3 p_{\min}^4} + \frac{\sigma^2 p_{\max}^{\frac{3}{2}} \vartheta \mu K^{\frac{3}{2}} \sqrt{T\log T}\max\{N,T\}}{\psi_{\min}^2 p_{\min}^3 \sqrt{N}}\right)$$

by Claims F.4 (ii) and F.7 (v). To bound the second term of $\delta_{1,i}$, define $\ddot{\gamma}_{i,s} = \frac{1}{N}\sum_{j=1}^N (\varepsilon_{is}\varepsilon_{js} - \mathbb{E}[\varepsilon_{is}\varepsilon_{js}|\Omega, \mathcal{M}])\omega_{is}\omega_{js}U_{M^\star, j}$. Then, conditioning on Ω and $\mathcal{M}$, for each i, $\{\ddot{\gamma}_{i,s}\}_{s\leq T}$ are independent across s. In addition, we have

$$\|\sum_{s=1}^T \mathbb{E}[\ddot{\gamma}_{i,s}'\ddot{\gamma}_{i,s}|\Omega, \mathcal{M}]\| \leq \frac{T}{N}\max_s \frac{1}{N}\sum_{j=1}^N\sum_{k=1}^N |\text{Cov}\left(\varepsilon_{is}\varepsilon_{js}, \varepsilon_{is}\varepsilon_{ks}|\Omega, \mathcal{M}\right)| \|U_{M^\star}\|_{2,\infty}^2 \leq C\frac{\sigma^2\mu K T}{N^2}.$$

Hence, by the matrix Bernstein inequality (ex. Koltchinskii et al. (2011) Proposition 2) with conditioning on Ω and $\mathcal{M}$, we have

$$\max_i \|\sum_{s=1}^T\sum_{j=1}^N (\varepsilon_{is}\varepsilon_{js} - \mathbb{E}[\varepsilon_{is}\varepsilon_{js}|\Omega, \mathcal{M}])\omega_{is}\omega_{js}H_1'\beta_j\| = N^{\frac{3}{2}}\max_i \|\sum_{s=1}^T \ddot{\gamma}_{i,s}\| = O_P\left(\sigma^2\mu^{\frac{1}{2}}K^{\frac{1}{2}}\max\{N\log^2 N, T\log^2 T\}\right).$$

Then, by Claims F.4 and F.7, the second term is bounded by $O_P\left(\frac{\sigma^2\mu^{\frac{1}{2}}K^{\frac{1}{2}}\max\{N\log^2 N, T\log^2 T\}}{\psi_{\min}^2 p_{\min}^2}\right)$. The third term can be bounded like

$$\frac{1}{NT}\max_i \|\widehat{A}_i^{-1}\|\|B^{-1}\| T \max_s \max_i \sum_{j=1}^N |\text{Cov}(\varepsilon_{is}, \varepsilon_{js}|\Omega, \mathcal{M})| \max_j \|H_1'\beta_j\| = O_P\left(\frac{\sigma^2\mu^{\frac{1}{2}}K^{\frac{1}{2}}T}{\psi_{\min}^2 p_{\min}^2}\right).$$

Step 2. Next, we bound $\max_i \|\delta_{2,i}\|$. Note that

$$\delta_{2,i} = -\frac{1}{NT}\widehat{A}_i^{-1}\sum_{s=1}^T \omega_{is}\widehat{F}_s\gamma_s'\widetilde{B}_s^{-1}H_2'^{-1}\beta_i$$
$$= \frac{1}{NT}\widehat{A}_i^{-1}\sum_{s=1}^T \omega_{is}H_2F_s\gamma_s'\left(B^{-1} - \widetilde{B}_s^{-1}\right)H_2'^{-1}\beta_i - \frac{1}{NT}\widehat{A}_i^{-1}\sum_{s=1}^T \omega_{is}\left(\widehat{F}_s - H_2F_s\right)\gamma_s'B^{-1}H_2'^{-1}\beta_i$$
$$- \frac{1}{NT}\widehat{A}_i^{-1}\sum_{s=1}^T \omega_{is}H_2F_s\gamma_s'B^{-1}H_2'^{-1}\beta_i + \frac{1}{NT}\widehat{A}_i^{-1}\sum_{s=1}^T \omega_{is}\left(\widehat{F}_s - H_2F_s\right)\gamma_s'\left(B^{-1} - \widetilde{B}_s^{-1}\right)H_2'^{-1}\beta_i.$$

Since the first three terms dominate the last term, we bound the first three terms. By Claims F.3 (iii) (iv), F.4 (ii), F.6 (iii) and F.7 (v), the first term is bounded by

$$O_P\left(\frac{\sigma^2 p_{\max}^2 c_{\text{inv}} q^3 \mu K^2 \sqrt{T}\max\{\sqrt{N}, \sqrt{T}\}}{\psi_{\min}^2 p_{\min}^4} + \frac{\sigma p_{\max}\vartheta\mu^{\frac{3}{2}}K^{\frac{5}{2}}\sqrt{T\log T}}{\psi_{\min}p_{\min}^3\sqrt{N}}\right).$$

Similarly, by Claim F.7 (iii), the second term can be bounded like

$$\frac{p_{\max}^{\frac{1}{2}}}{N\sqrt{T}}\max_i \|\widehat{A}_i^{-1}\|\left(\sum_{s=1}^T \|\widehat{F}_s - H_2F_s\|^2\right)^{\frac{1}{2}}\max_s \|\gamma_s\|\|B^{-1}\|\max_i \|H_2'^{-1}\beta_i\| = O_P\left(\frac{\sigma^2 p_{\max}\mu K^{\frac{3}{2}} T\sqrt{\log T}}{\psi_{\min}^2 p_{\min}^3}\right)$$

Because $\sum_{s=1}^T \omega_{is}H_2F_s\gamma_s' = (I_K + \varphi)\sum_{s=1}^T \omega_{is}H_1^{-1}F_s\gamma_s'$, to bound the third term, we need to estimate $\max_i \|\sum_{s=1}^T \omega_{is}H_1^{-1}F_s\gamma_s'\|$. Note that $\|\sum_{s=1}^T \omega_{is}H_1^{-1}F_s\gamma_s'\| = \|\sum_{s=1}^T\sum_{j=1}^N \varepsilon_{js}\omega_{js}\omega_{is}H_1^{-1}F_s\beta_j'H_1\|$. Conditioning on $\mathcal{M}$ and Ω, $\{\varepsilon_{js}\}_{j\leq N, s\leq T}$ are independent. Hence, by the matrix Bernstein inequality with conditioning on $\mathcal{M}$ and Ω, we have $\max_i \|\sum_{s=1}^T\sum_{j=1}^N \varepsilon_{js}\omega_{js}\omega_{is}H_1^{-1}F_s\beta_j'H_1\|$

$O_P\left(\sigma qK\sqrt{\log N}\psi_{\min}\right)$. Hence, by Claims F.3, F.4 and F.7, the third term is bounded like

$$\frac{1}{NT}\max_i \|\widehat{A}_i^{-1}\|\|I_K+\varphi\|\max_i\|\sum_{s=1}^T \omega_{is}H_1^{-1}F_s\gamma_s'\|\|B^{-1}\|\max_i\|H_2'^{-1}\beta_i\| = O_P\left(\frac{\sigma q\mu^{\frac{1}{2}}K^{\frac{3}{2}}\sqrt{\log N}}{p_{\min}^2\psi_{\min}}\right).$$

Step 3. We bound $\max_i\|\delta_{3,i}\|$. Note that

$$\delta_{3,i} = -\sum_{d=1}^4 \widehat{A}_i^{-1}\frac{1}{T}\sum_{s=1}^T\omega_{is}(\widehat{F}_s - H_2F_s)\Delta_{d,s}'H_2'^{-1}\beta_i - \sum_{d=1}^4\widehat{A}_i^{-1}\frac{1}{T}\sum_{s=1}^T\omega_{is}H_2F_s\Delta_{d,s}'H_2'^{-1}\beta_i.$$

The first term can be bounded by Claims F.3 and F.7 like

$$\max_i\|\widehat{A}_i^{-1}\|\left(\frac{1}{T}\sum_{s=1}^T\|\widehat{F}_s-H_2F_s\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^T\omega_{is}\right)^{\frac{1}{2}}\max_s\|R_s^F\|\max_i\|H_2'^{-1}\beta_i\| = O_P\left(\frac{\sigma p_{\max}\mu K\sqrt{NT}}{\psi_{\min}^2p_{\min}^2}\right)\max_s\|R_s^F\|$$

In addition, the second term is bounded like

$$\frac{1}{T}\max_i\|\widehat{A}_i^{-1}\|\|H_2F'\|\sqrt{T}\max_s\|R_s^F\|\max_i\|H_2'^{-1}\beta_i\| = O_P\left(\frac{p_{\max}q\mu K\sqrt{NT}}{\psi_{\min}p_{\min}}\right)\max_s\|R_s^F\|.$$

Step 4. Note that $\max_i\|\delta_{4,i}\| \le \sum_{d=1}^4\frac{1}{T}\max_i\|\widehat{A}_i^{-1}\|\sqrt{T}\max_s\|\Delta_{d,s}\|\max_i\|\mathcal{E}_i\|$, $\max_i\|\mathcal{E}_i\| = \max_i\sqrt{\sum_{s=1}^T\varepsilon_{is}^2} = O_P(\sigma\sqrt{T})$.
Then, we have $\max_i\|\delta_{4,i}\| \le \frac{1}{T}\max_i\|\widehat{A}_i^{-1}\|\sqrt{T}\max_s\|R_s^F\|\max_i\|\mathcal{E}_i\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}}NT}{\psi_{\min}^2p_{\min}}\right)\max_s\|R_s^F\|$.
Step 5. We bound $\max_i\|\delta_{5,i}\|$. By Claims F.6 and F.7, we have

$$\max_i\|\delta_{5,i}\| \le \max_i\|\widehat{A}_i^{-1}-A_i^{-1}\|\max_i\|\frac{1}{T}\sum_{s=1}^T\omega_{is}H_2F_s\varepsilon_{is}\| = O_P\left(\frac{\sigma^2p_{\max}^{\frac{3}{2}}q^2\mu^{\frac{1}{2}}K^{\frac{3}{2}}\sqrt{NT\log N}}{p_{\min}^3\psi_{\min}^2}\right).$$

Step 6. Finally, we bound $\max_i\|\delta_{6,i}\|$. Note that $\frac{1}{T}\sum_{s=1}^T\|\widehat{F}_s\| \le \frac{1}{T}\sum_{s=1}^T\|\widehat{F}_s-H_2F_s\|+\frac{1}{T}\sum_{s=1}^T\|H_2F_s\| = O_P\left(\frac{qK^{\frac{1}{2}}\psi_{\min}}{\sqrt{NT}}\right)$.
So, we have $\max_i\|\widehat{A}_i^{-1}\|\max_{is}|M_{is}^R|\frac{1}{T}\sum_{s=1}^T\|\widehat{F}_s\| = O_P\left(\frac{NT}{\psi_{\min}^2}\frac{qK^{\frac{1}{2}}\psi_{\min}}{\sqrt{NT}}\right)\max_{it}|M_{it}^R|$ by Claims F.3 and F.7.
Then, by using the bound for $\max_s\|R_s^F\|$ in Proposition E.1, we can get the desired result. □

F.2 Proofs for Section 3

F.2.1 Proof of Theorem 3.1

From Lemma E.4, we can know that Assumptions C.1 - C.3 with $\mu = C\eta$ for some constant $C>0$ are satisfied under the assumptions of Theorem 3.1. Hence, we have Theorem 3.1 from Theorem C.1 and Lemma E.4.

Proof of Theorem C.1

To begin with, we define a few more notations to facilitate the proofs. Let $\mathcal{E}_t = [\varepsilon_{1t},\dots,\varepsilon_{Nt}]'$, $\mathcal{E}_i = [\varepsilon_{i1},\dots,\varepsilon_{iT}]'$, $M_t^R = [M_{1t}^R,\dots,M_{Nt}^R]'$ and $M_i^R = [M_{i1}^R,\dots,M_{iT}^R]'$. Also, let $\Omega_t = \mathrm{diag}(\omega_{1t},\dots,\omega_{Nt})$ and $\Omega_i = \mathrm{diag}(\omega_{i1},\dots,\omega_{iT})$.
By Propositions E.1 and F.1, we can have following decomposition:

$$\mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}}\left(\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\widehat{\beta}_i'\widehat{F}_t - \frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}M_{it}\right) = \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\left(\widehat{\beta}_i'\widehat{F}_t - \beta_i'H_2^{-1}H_2F_t\right) - \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}M_{it}^R$$

$$= \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\left(\beta_i'\left(\sum_{j=1}^N\omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N\omega_{jt}\beta_j\varepsilon_{jt}\right) + F_t'\left(\sum_{s=1}^T\omega_{is}F_sF_s'\right)^{-1}\left(\sum_{s=1}^T\omega_{is}F_s\varepsilon_{is}\right)\right) + \sum_{d=1}^7\mathcal{R}_{d,\mathcal{G}},$$

$$\mathcal{R}_{1,\mathcal{G}} \coloneqq \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o} \sum_{(i,t)\in\mathcal{G}} \beta_i' H_2^{-1} R_t^F, \quad \mathcal{R}_{2,\mathcal{G}} \coloneqq \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o} \sum_{(i,t)\in\mathcal{G}} R_i^{\beta\prime} H_2 F_t,$$

$$\mathcal{R}_{3,\mathcal{G}} \coloneqq \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o} \sum_{(i,t)\in\mathcal{G}} \left(\sum_{s=1}^T \omega_{is} F_s \varepsilon_{is}\right)' \left(\sum_{s=1}^T \omega_{is} F_s F_s'\right)^{-1} \left(\sum_{j=1}^N \omega_{jt} \beta_j \beta_j'\right)^{-1} \left(\sum_{j=1}^N \omega_{jt} \beta_j \varepsilon_{jt}\right),$$

$$\mathcal{R}_{4,\mathcal{G}} \coloneqq \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o} \sum_{(i,t)\in\mathcal{G}} \left(\sum_{s=1}^T \omega_{is} F_s \varepsilon_{is}\right)' \left(\sum_{s=1}^T \omega_{is} F_s F_s'\right)^{-1} H_2^{-1} R_t^F,$$

$$\mathcal{R}_{5,\mathcal{G}} \coloneqq \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o} \sum_{(i,t)\in\mathcal{G}} R_i^{\beta\prime} H_2 \left(\sum_{j=1}^N \omega_{jt} \beta_j \beta_j'\right)^{-1} \left(\sum_{j=1}^N \omega_{jt} \beta_j \varepsilon_{jt}\right),$$

$$\mathcal{R}_{6,\mathcal{G}} \coloneqq \mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o} \sum_{(i,t)\in\mathcal{G}} R_i^{\beta\prime} R_t^F, \quad \mathcal{R}_{7,\mathcal{G}} \coloneqq -\mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o} \sum_{(i,t)\in\mathcal{G}} M_{it}^R.$$

Step 1. To show $\sum_{d=1}^7 |\mathcal{R}_{d,\mathcal{G}}| = o_P(1)$, we bound $\mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}}$ first. Note that

$$\begin{aligned}
&\frac{1}{|\mathcal{T}|_o^2} \sum_{t\in\mathcal{T}} \bar{\beta}_{\mathcal{I}}' \left(\sum_{j=1}^N \omega_{jt} \beta_j \beta_j'\right)^{-1} \left(\sum_{j=1}^N \omega_{jt} \sigma_{jt}^2 \beta_j \beta_j'\right) \left(\sum_{j=1}^N \omega_{jt} \beta_j \beta_j'\right)^{-1} \bar{\beta}_{\mathcal{I}} \\
&= \frac{1}{|\mathcal{T}|_o^2} \sum_{t\in\mathcal{T}} \bar{U}_{\mathcal{I}}' \left(\sum_{j=1}^N \omega_{jt} U_j U_j'\right)^{-1} \left(\sum_{j=1}^N \omega_{jt} \sigma_{jt}^2 U_j U_j'\right) \left(\sum_{j=1}^N \omega_{jt} U_j U_j'\right)^{-1} \bar{U}_{\mathcal{I}} \\
&\geq \frac{1}{|\mathcal{T}|_o^2} \sum_{t\in\mathcal{T}} \left\|\bar{U}_{\mathcal{I}}\right\|^2 \psi_{\min}\left(\left(\sum_{j=1}^N \omega_{jt} U_j U_j'\right)^{-1}\right)^2 \psi_{\min}\left(\sum_{j=1}^N \omega_{jt} \sigma_{jt}^2 U_j U_j'\right).
\end{aligned}$$

We have with probability converging to 1,

$$\psi_{\min}\left(\sum_{j=1}^N \omega_{jt} \sigma_{jt}^2 U_j U_j'\right) \geq \psi_{\min}\left(\sum_{j=1}^N p_j \sigma_{jt}^2 U_j U_j'\right) - \left\|\sum_{j=1}^N (\omega_{jt} - p_j) \sigma_{jt}^2 U_j U_j'\right\| \geq \frac{p_{\min}\sigma_{\min}^2}{2}$$

since $\psi_{\min}(\sum_{j=1}^N p_j \sigma_{jt}^2 U_j U_j') = \psi_{\min}(U'\Omega_{p\sigma}^* U) \geq \psi_{\min}(U)^2 \psi_{\min}(\Omega_{p\sigma}^*) \geq p_{\min}\sigma_{\min}^2$ and $\left\|\sum_{j=1}^N (\omega_{jt} - p_j) \sigma_{jt}^2 U_j U_j'\right\| \leq \frac{p_{\min}\sigma_{\min}^2}{2}$. Here, $\Omega_{p\sigma}^*$ is the $N \times N$ diagonal matrix of $\{p_j \sigma_{jt}^2\}_{j\leq N}$. In addition, $\psi_{\min}\left(\left(\sum_{j=1}^N \omega_{jt} U_j U_j'\right)^{-1}\right) \geq \frac{1}{p_{\max}}$ since

$$\left\|\sum_{j=1}^N \omega_{jt} U_j U_j'\right\| \leq \left\|\sum_{j=1}^N p_j U_j U_j'\right\| + \left\|\sum_{j=1}^N (\omega_{jt} - p_j) U_j U_j'\right\| = \left\|U'\Omega_p^* U\right\| + \left\|\sum_{j=1}^N (\omega_{jt} - p_j) U_j U_j'\right\| \lesssim p_{\max}$$

where Ω_p^* is the $N \times N$ diagonal matrix of $\{p_j\}_{j\leq N}$. Hence, we have

$$\frac{1}{|\mathcal{T}|_o^2} \sum_{t\in\mathcal{T}} \bar{\beta}_{\mathcal{I}}' \left(\sum_{j=1}^N \omega_{jt} \beta_j \beta_j'\right)^{-1} \left(\sum_{j=1}^N \omega_{jt} \sigma_{jt}^2 \beta_j \beta_j'\right) \left(\sum_{j=1}^N \omega_{jt} \beta_j \beta_j'\right)^{-1} \bar{\beta}_{\mathcal{I}} \geq c\sigma_{\min}^2 \frac{p_{\min}}{p_{\max}^2} \frac{1}{N|\mathcal{T}|_o}.$$

Similarly, we have

$$\frac{1}{|\mathcal{I}|_o^2} \sum_{i\in\mathcal{I}} \bar{F}_{\mathcal{T}}' \left(\sum_{s=1}^T \omega_{is} F_s F_s'\right)^{-1} \left(\sum_{s=1}^T \omega_{is} \sigma_{is}^2 F_s F_s'\right) \left(\sum_{s=1}^T \omega_{is} F_s F_s'\right)^{-1} \bar{F}_{\mathcal{T}} \geq c\sigma_{\min}^2 \frac{p_{\min}}{p_{\max}^2} \frac{1}{T|\mathcal{I}|_o}.$$

Hence, we have $\mathcal{V}_\mathcal{G} \geq c \frac{\sigma^2_{\min}}{\min\{|\mathcal{T}|_o N, |\mathcal{I}|_o T\}} \frac{p_{\min}}{p^2_{\max}}$ for some constant $c > 0$ and

$$\mathcal{V}_\mathcal{G}^{-\frac{1}{2}} = O_P\left(\frac{\min\{|\mathcal{I}|_o^{1/2}, |\mathcal{T}|_o^{1/2}\}\max\{\sqrt{N}, \sqrt{T}\}}{\sigma_{\min}} \frac{p_{\max}}{p_{\min}^{1/2}}\right).$$

Step 2. We now show $\sum_{d=1}^7 |\mathcal{R}_{d,\mathcal{G}}| = o_P(1)$. We have by Proposition E.1 and Claim F.3

$$|\mathcal{R}_{1,\mathcal{G}}| \leq \mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \max_i \|\beta_i' H_2^{-1}\| \max_t \|R_t^F\| = o_P(1).$$

Similarly, by Proposition F.1 and Claim F.3, we obtain $|\mathcal{R}_{2,\mathcal{G}}| \leq \mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \max_i \|R_i^{\beta\prime}\| \max_t \|H_2 F_t\| = o_P(1)$. By Claims F.4, F.6 and F.7, we have

$$|\mathcal{R}_{3,\mathcal{G}}| \leq \mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \max_i \left\| H_2'^{-1}\left(\sum_{s=1}^T \omega_{is} F_s F_s'\right)^{-1} H_2^{-1}\right\| \max_t \left\| H_2 \left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1} H_2'\right\| \frac{1}{|\mathcal{I}|_o}\sum_{i\in\mathcal{I}}\left\|\sum_{s=1}^T \omega_{is} H_2 F_s \varepsilon_{is}\right\|$$
$$\times \frac{1}{|\mathcal{T}|_o}\sum_{t\in\mathcal{T}}\left\|\sum_{j=1}^N \omega_{jt} H_2'^{-1}\beta_j\varepsilon_{jt}\right\| = O_P\left(\frac{\sigma p^2_{\max}\min\{|\mathcal{I}|_o^{1/2}, |\mathcal{T}|_o^{1/2}\} qK\sqrt{\log N}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}^{5/2}\psi_{\min}}\right) = o_P(1).$$

Moreover, by Proposition E.1 with Claims F.6 and F.7,

$$|\mathcal{R}_{4,\mathcal{G}}| \leq \mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \max_i \left\|\sum_{s=1}^T \omega_{is} H_2 F_s \varepsilon_{is}\right\| \max_i \left\| H_2'^{-1}\left(\sum_{s=1}^T \omega_{is} F_s F_s'\right)^{-1} H_2^{-1}\right\| \max_t \|R_t^F\| = o_P(1).$$

By Proposition F.1, $|\mathcal{R}_{5,\mathcal{G}}| \leq \mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \max_i \|R_i^\beta\| \max_t \left\| H_2\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N \omega_{jt}\beta_j\varepsilon_{jt}\right)\right\| = o_P(1)$. In addition, we have by Propositions E.1 and F.1 that $|\mathcal{R}_{6,\mathcal{G}}| \leq \mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \max_i \|R_i^\beta\| \max_t \|R_t^F\| = o_P(1)$. Lastly, we have by Assumption C.3 that $|\mathcal{R}_{7,\mathcal{G}}| = \left|\mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}} M_{it}^R\right| \leq \mathcal{V}_\mathcal{G}^{-\frac{1}{2}} \max_{it}|M_{it}^R| = o_P(1)$. Therefore, we obtain $\sum_{d=1}^7 |\mathcal{R}_{d,\mathcal{G}}| = o_P(1)$. □

Step 3. Finally, we show $\mathcal{V}_\mathcal{G}^{-\frac{1}{2}}\mathcal{H} \xrightarrow{D} \mathcal{N}(0,1)$ where

$$\mathcal{H} := \frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\left(\beta_i'\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N \omega_{jt}\beta_j\varepsilon_{jt}\right) + F_t'\left(\sum_{s=1}^T \omega_{is}F_sF_s'\right)^{-1}\left(\sum_{s=1}^T \omega_{is}F_s\varepsilon_{is}\right)\right).$$

Define

$$\Phi_{1,NT} := \mathcal{V}_{1,NT}^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\beta_i'\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\sum_{j=1}^N \omega_{jt}\beta_j\varepsilon_{jt}, \quad \Phi_{2,NT} := \mathcal{V}_{2,NT}^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}F_t'\left(\sum_{s=1}^T \omega_{is}F_sF_s'\right)^{-1}\sum_{s=1}^T \omega_{is}F_s\varepsilon_{is},$$
$$\mathcal{V}_{1,NT} := \frac{1}{|\mathcal{T}|_o^2}\sum_{t\in\mathcal{T}}\bar{\beta}_\mathcal{I}'\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N \omega_{jt}\sigma_{jt}^2\beta_j\beta_j'\right)\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\bar{\beta}_\mathcal{I}$$
$$\mathcal{V}_{2,NT} := \frac{1}{|\mathcal{I}|_o^2}\sum_{i\in\mathcal{I}}\bar{F}_\mathcal{T}'\left(\sum_{s=1}^T \omega_{is}F_sF_s'\right)^{-1}\left(\sum_{s=1}^T \omega_{is}\sigma_{is}^2F_sF_s'\right)\left(\sum_{s=1}^T \omega_{is}F_sF_s'\right)^{-1}\bar{F}_\mathcal{T}.$$

We can show that, conditioning on $\{\mathcal{M},\Omega\}$, $(\Phi_{1,NT},\Phi_{2,NT})$ converges (jointly) to a bivariate normal distribution. Let

$\Xi_{1,t} \coloneqq \sum_{j=1}^{N} \omega_{jt}\beta_j\beta_j'$, $\Xi_{2,i} \coloneqq \sum_{s=1}^{T} \omega_{is}F_sF_s'$. Note that for any constants a, b, we have

$$a\Phi_{1,NT} + b\Phi_{2,NT} = \sum_{j=1}^{N}\sum_{s=1}^{T} \varepsilon_{js}\left(a\omega_{js}\frac{1}{|\mathcal{T}|_o}\mathcal{V}_{1,NT}^{-\frac{1}{2}}\bar{\beta}_{\mathcal{I}}'\Xi_{1,s}^{-1}\beta_j 1\{s\in\mathcal{T}\} + b\omega_{js}\frac{1}{|\mathcal{I}|_o}\mathcal{V}_{2,NT}^{-\frac{1}{2}}\bar{F}_{\mathcal{T}}'\Xi_{2,j}^{-1}F_s 1\{j\in\mathcal{I}\}\right).$$

In addition, a simple calculation shows that

$$\mathrm{Var}\left(a\Phi_{1,NT} + b\Phi_{2,NT}|\mathcal{M},\Omega\right) = a^2 + b^2 + \frac{2ab}{|\mathcal{I}|_o|\mathcal{T}|_o}\mathcal{V}_{1,NT}^{-\frac{1}{2}}\mathcal{V}_{2,NT}^{-\frac{1}{2}}\sum_{j\in\mathcal{I}}\sum_{s\in\mathcal{T}}\omega_{js}\sigma_{js}^2\bar{\beta}_{\mathcal{I}}'\Xi_{1,s}^{-1}\beta_j\bar{F}_{\mathcal{T}}'\Xi_{2,j}^{-1}F_s$$

and the last term is $o_P(1)$ by Claims F.3, F.4 and F.7. Hence, we have $\mathrm{Var}\left(a\Phi_{1,NT} + b\Phi_{2,NT}|\mathcal{M},\Omega\right) \xrightarrow{P} a^2 + b^2$. Next, we check the Lindeberg's condition. First, note that $\max_{j,t}|\bar{\beta}_{\mathcal{I}}'\Xi_{1,t}^{-1}\beta_j| = O_P\left(\frac{\mu K}{p_{\min}N}\right)$ by Claims F.3 and F.4. Let $\mathcal{Y}_{1,jt} \coloneqq \frac{1}{|\mathcal{T}|_o}\mathcal{V}_{1,NT}^{-\frac{1}{2}}\omega_{jt}\varepsilon_{jt}\bar{\beta}_{\mathcal{I}}'\Xi_{1,t}^{-1}\beta_j$. Then, we have

$$\sum_{t\in\mathcal{T}}\sum_{j=1}^{N}\mathbb{E}\left[\mathcal{Y}_{1,jt}^4|\mathcal{M},\Omega\right] = \frac{1}{|\mathcal{T}|_o^4}\mathcal{V}_{1,NT}^{-2}\sum_{t\in\mathcal{T}}\sum_{j=1}^{N}\mathbb{E}[\varepsilon_{jt}^4|\mathcal{M},\Omega]\omega_{jt}\left(\beta_j'\Xi_{1,t}^{-1}\bar{\beta}_{\mathcal{I}}\right)^4 = O_P\left(\frac{p_{\max}^3\mu^4K^4}{p_{\min}^5|\mathcal{T}|_oN}\right),$$

because $\mathcal{V}_{1,NT}^{-\frac{1}{2}} = O_P\left(\frac{p_{\max}\mu K|\mathcal{T}|_o^{1/2}\sqrt{N}}{p_{\min}^{1/2}\sigma_{\min}}\right)$. Similarly, we have $\sum_{i\in\mathcal{I}}\sum_{s=1}^{T}\mathbb{E}\left[\mathcal{Y}_{2,is}^4|\mathcal{M},\Omega\right] = O_P\left(\frac{p_{\max}^3\mu^4K^4}{p_{\min}^5|\mathcal{I}|_oT}\right)$ where $\mathcal{Y}_{2,is} \coloneqq \frac{1}{|\mathcal{I}|_o}\mathcal{V}_{2,NT}^{-\frac{1}{2}}\omega_{is}\varepsilon_{is}\bar{F}_{\mathcal{T}}'\Xi_{2,i}^{-1}F_s$. Then,

$$\begin{aligned}&\sum_{j=1}^{N}\sum_{s=1}^{T}\mathbb{E}\left[\varepsilon_{js}^4\left(a\omega_{js}\frac{1}{|\mathcal{T}|_o}\mathcal{V}_{1,NT}^{-\frac{1}{2}}\bar{\beta}_{\mathcal{I}}'\Xi_{1,s}^{-1}\beta_j 1\{s\in\mathcal{T}\} + b\omega_{js}\frac{1}{|\mathcal{I}|_o}\mathcal{V}_{2,NT}^{-\frac{1}{2}}\bar{F}_{\mathcal{T}}'\Xi_{2,j}^{-1}F_s 1\{j\in\mathcal{I}\}\right)^4\middle|\mathcal{M},\Omega\right]\\&\le C\left(\sum_{s\in\mathcal{T}}\sum_{j=1}^{N}a^4\mathbb{E}\left[\mathcal{Y}_{1,js}^4|\mathcal{M},\Omega\right] + \sum_{j\in\mathcal{I}}\sum_{s=1}^{T}b^4\mathbb{E}\left[\mathcal{Y}_{2,js}^4|\mathcal{M},\Omega\right]\right) = O_P\left(\frac{p_{\max}^3\mu^4K^4}{p_{\min}^5N} + \frac{p_{\max}^3\mu^4K^4}{p_{\min}^5T}\right).\end{aligned}$$

Let $\mathcal{Y}_{3,js} \coloneqq \varepsilon_{js}\left(a\omega_{js}\frac{1}{|\mathcal{T}|_o}\mathcal{V}_{1,NT}^{-\frac{1}{2}}\bar{\beta}_{\mathcal{I}}'\Xi_{1,s}^{-1}\beta_j 1\{s\in\mathcal{T}\} + b\omega_{js}\frac{1}{|\mathcal{I}|_o}\mathcal{V}_{2,NT}^{-\frac{1}{2}}\bar{F}_{\mathcal{T}}'\Xi_{2,j}^{-1}F_s 1\{j\in\mathcal{I}\}\right)$. Then, for all $\epsilon > 0$, we have the following Lindeberg's condition:

$$\begin{aligned}&\mathrm{Var}\left(a\Phi_{1,NT} + b\Phi_{2,NT}|\mathcal{M},\Omega\right)^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}\mathbb{E}\left[\mathcal{Y}_{3,js}^2 1\left\{|\mathcal{Y}_{3,js}| > \epsilon\mathrm{Var}\left(a\Phi_{1,NT} + b\Phi_{2,NT}|\mathcal{M},\Omega\right)^{1/2}\right\}|\mathcal{M},\Omega\right]\\&\le \frac{1}{\mathrm{Var}\left(a\Phi_{1,NT} + b\Phi_{2,NT}|\mathcal{M},\Omega\right)\epsilon}\sqrt{\sum_{j=1}^{N}\sum_{s=1}^{T}\mathbb{E}\left[\mathcal{Y}_{3,js}^4|\mathcal{M},\Omega\right]} = O_P\left(\frac{p_{\max}^{3/2}}{p_{\min}^{5/2}}\sqrt{\frac{\mu^4K^4}{N} + \frac{\mu^4K^4}{T}}\right) = o_P(1).\end{aligned}$$

Hence, by the central limit theorem, we have, conditioning on $\{\mathcal{M},\Omega\}$, $a\Phi_{1,NT} + b\Phi_{2,NT} \xrightarrow{D} \mathcal{N}\left(0, a^2+b^2\right)$, and it means that, for any constants a, b, conditioning on $\{\mathcal{M},\Omega\}$, we have $a\Phi_{1,NT} + b\Phi_{2,NT} \xrightarrow{D} a\Phi_1 + b\Phi_2$, where $\begin{bmatrix}\Phi_1\\ \Phi_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)$. So, by the Cramer-Wold theorem, we can say that conditioning on $\{\mathcal{M},\Omega\}$, $(\Phi_{1,NT}, \Phi_{2,NT})$ converges (jointly) to the bivariate normal random vector (Φ_1, Φ_2). Hence, by using the same argument as in the proof of Theorem 3 in Bai (2003), we have, conditioning on $\{\mathcal{M},\Omega\}$, $\mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}}\mathcal{H} = \frac{\mathcal{A}_{NT}\Phi_{1,NT} + \mathcal{B}_{NT}\Phi_{2,NT}}{\sqrt{\mathcal{A}_{NT}^2\mathrm{Var}(\Phi_1|\mathcal{M},\Omega) + \mathcal{B}_{NT}^2\mathrm{Var}(\Phi_2|\mathcal{M},\Omega)}} \xrightarrow{D} \mathcal{N}(0,1)$ where $\mathcal{A}_{NT} \coloneqq \frac{\mathcal{V}_{1,NT}^{1/2}}{\max\left\{\mathcal{V}_{1,NT}^{1/2}, \mathcal{V}_{2,NT}^{1/2}\right\}}$ and $\mathcal{B}_{NT} \coloneqq \frac{\mathcal{V}_{2,NT}^{1/2}}{\max\left\{\mathcal{V}_{1,NT}^{1/2}, \mathcal{V}_{2,NT}^{1/2}\right\}}$. Lastly, the unconditional convergence is due to the dominated convergence theorem. □

F.2.2 Proof of Theorem 3.2

The result follows from Theorem C.2. We now prove this theorem.

Proof of Theorem C.2.

Thanks to Theorem C.1, it suffices to show $\frac{\mathcal{V}_{\mathcal{G}}-\widehat{\mathcal{V}}_{\mathcal{G}}}{\mathcal{V}_{\mathcal{G}}} = o_P(1)$ instead.

$$\begin{aligned}\mathcal{V}_{\mathcal{G}}^{-1}|\widehat{\mathcal{V}}_{\mathcal{G}} - \mathcal{V}_{\mathcal{G}}| \leq \mathcal{V}_{\mathcal{G}}^{-1}\frac{1}{|\mathcal{T}|_o}\max_t\Bigg|&\widehat{\bar{\beta}}_{\mathcal{I}}'\left(\sum_{j=1}^N \omega_{jt}\widehat{\beta}_j\widehat{\beta}_j{}'\right)^{-1}\left(\sum_{j=1}^N \omega_{jt}\widehat{\sigma}_j^2\widehat{\beta}_j\widehat{\beta}_j{}'\right)\left(\sum_{j=1}^N \omega_{jt}\widehat{\beta}_j\widehat{\beta}_j{}'\right)^{-1}\widehat{\bar{\beta}}_{\mathcal{I}}\\ &-\bar{\beta}_{\mathcal{I}}'\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N \omega_{jt}\sigma_j^2\beta_j\beta_j'\right)\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\bar{\beta}_{\mathcal{I}}\Bigg| \quad \cdots\cdots \quad (a)\\ +\mathcal{V}_{\mathcal{G}}^{-1}\frac{1}{|\mathcal{I}|_o}\max_i\Bigg|&\widehat{\sigma}_i^2\widehat{\bar{F}}_{\mathcal{T}}'\left(\sum_{s=1}^T \omega_{is}\widehat{F}_s\widehat{F}_s'\right)^{-1}\widehat{\bar{F}}_{\mathcal{T}} - \sigma_i^2\bar{F}_{\mathcal{T}}'\left(\sum_{s=1}^T \omega_{is}F_sF_s'\right)^{-1}\bar{F}_{\mathcal{T}}\Bigg| \quad \cdots\cdots \quad (b).\end{aligned}$$

In addition, as noted in Step 1 of the proof of Theorem C.1, we have $\mathcal{V}_{\mathcal{G}}^{-1}\frac{1}{|\mathcal{I}|_o} = O_P(\frac{p_{\max}^2\max\{N,T\}}{p_{\min}\sigma_{\min}^2})$, $\mathcal{V}_{\mathcal{G}}^{-1}\frac{1}{|\mathcal{T}|_o} = O_P(\frac{p_{\max}^2\max\{N,T\}}{p_{\min}\sigma_{\min}^2})$. Then, we obtain

$$\begin{aligned}(a) = {}& O_P\left(\frac{p_{\max}^2\max\{N,T\}}{p_{\min}\sigma_{\min}^2}\right)\|\widehat{\bar{\beta}}_{\mathcal{I}}' - \bar{\beta}_{\mathcal{I}}'H_2^{-1}\|\max_t\left\|\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N \omega_{jt}\sigma_j^2\beta_j\beta_j'\right)\left(\sum_{j=1}^N \omega_{jt}\beta_j\beta_j'\right)^{-1}\bar{\beta}_{\mathcal{I}}\right\|\\ &+O_P\left(\frac{p_{\max}^2\max\{N,T\}}{p_{\min}\sigma_{\min}^2}\right)\|\bar{\beta}_{\mathcal{I}}'H_2^{-1}\|^2\max_t\left\|\left(\sum_{j=1}^N \omega_{jt}\widehat{\beta}_j\widehat{\beta}_j'\right)^{-1} - \left(\sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right)^{-1}\right\|\\ &\times\max_t\left\|\left(\sum_{j=1}^N \omega_{jt}\sigma_j^2H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right)\left(\sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right)^{-1}\right\|\\ &+O_P\left(\frac{p_{\max}^2\max\{N,T\}}{p_{\min}\sigma_{\min}^2}\right)\|\bar{\beta}_{\mathcal{I}}'H_2^{-1}\|^2\max_t\left\|\left(\sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right)^{-1}\right\|^2\max_t\left\|\sum_{j=1}^N \omega_{jt}\widehat{\sigma}_j^2\widehat{\beta}_j\widehat{\beta}_j' - \sum_{j=1}^N \omega_{jt}\sigma_j^2H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right\|.\end{aligned}$$

By Claims F.3, F.4 and F.8, the above three terms are $o_p(1)$. Hence, we have $(a) = o_P(1)$. Similarly, we can show that $(b) = o_p(1)$. This completes the proof. □

F.2.3 Proof of Theorem 3.3

Let (M, p, σ) be a point in $\mathcal{A}$. Without loss of generality, assume that $\mathcal{I}$ and $\mathcal{T}$ consist of the first elements of $\{1, \ldots, N\}$ and $\{1, \ldots, T\}$, respectively. We then partition $\beta = \begin{pmatrix}\beta_{\mathcal{I}}\\ \beta_{-\mathcal{I}}\end{pmatrix}$ with $\beta_{\mathcal{I}}$ being the first $|\mathcal{I}|_o$ rows of β. Similarly, let $F = \begin{pmatrix}F_{\mathcal{T}}\\ F_{-\mathcal{T}}\end{pmatrix}$. For notational simplicity, we denote $\chi_1 = \frac{\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}}}{|\mathcal{T}|_o}$, and $\chi_2 = \frac{\bar{F}_{\mathcal{T}}'(F'F)^{-1}\bar{F}_{\mathcal{T}}}{|\mathcal{I}|_o}$. For a fixed but arbitrary constant $\alpha \in (0, \frac{\sigma}{4\sqrt{p}})$, we define $\breve{M} = \breve{M}^\star + M^R$ such that $\breve{M}^\star = \breve{\beta}\breve{F}'$ with $\breve{\beta} = \begin{pmatrix}\breve{\beta}_{\mathcal{I}}\\ \beta_{-\mathcal{I}}\end{pmatrix}$ with $\breve{\beta}_i = \beta_i + \frac{\alpha}{\sqrt{\chi_1+\chi_2}}\frac{(F'F)^{-1}\bar{F}_{\mathcal{T}}}{|\mathcal{I}|_o}$ for each $i = 1, \ldots, |\mathcal{I}|_o$, and $\breve{F} = \begin{pmatrix}\breve{F}_{\mathcal{T}}\\ F_{-\mathcal{T}}\end{pmatrix}$ with $\breve{F}_t = F_t + \frac{\alpha}{\sqrt{\chi_1+\chi_2}}\frac{(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}}}{|\mathcal{T}|_o}$ for each $t = 1, \ldots, |\mathcal{T}|_o$. We define $s_*^2(\breve{M}, p, \sigma) = \frac{\sigma^2}{p}\frac{1}{|\mathcal{T}|_o}\bar{\breve{\beta}}_{\mathcal{I}}'(\breve{\beta}'\breve{\beta})^{-1}\bar{\breve{\beta}}_{\mathcal{I}} + \frac{\sigma^2}{p}\frac{1}{|\mathcal{I}|_o}\bar{\breve{F}}_{\mathcal{T}}'(\breve{F}'\breve{F})^{-1}\bar{\breve{F}}_{\mathcal{T}}$ where $\bar{\breve{\beta}}_{\mathcal{I}} = \frac{1}{|\mathcal{I}|_o}\sum_{i\in\mathcal{I}}\breve{\beta}_i$ and $\bar{\breve{F}}_{\mathcal{T}} = \frac{1}{|\mathcal{T}|_o}\sum_{t\in\mathcal{T}}\breve{F}_t$. We will proceed conditioning on the event $\psi_{\min}^{-1}(S_\beta) < \eta$ and $\psi_{\min}^{-1}(S_F) < \eta$. Then by the asymptotic unbiasedness, Claim F.9 (i), we have

$$\mathbb{E}_{\breve{M},p,\sigma}U(Y,\Omega) - \mathbb{E}_{M,p,\sigma}U(Y,\Omega) = \alpha\sqrt{\chi_1+\chi_2} + \frac{\alpha^2}{(\chi_1+\chi_2)|\mathcal{I}|_o|\mathcal{T}|_o}\bar{F}_{\mathcal{T}}'(F'F)^{-1}(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}} + o(s_*(M,p,\sigma)) + o(s_*(\breve{M},p,\sigma))$$

$$= \alpha\sqrt{\chi_1+\chi_2} + o(s_*(M,p,\sigma)).$$

On the other hand, by Cauchy-Schwarz inequality,

$$\mathbb{E}_{\breve{M},p,\sigma}\left[U(Y,\Omega)\right] - \mathbb{E}_{M,p,\sigma}\left[U(Y,\Omega)\right] = \mathbb{E}_{M,p,\sigma}\left[\left(U(Y,\Omega) - |\mathcal{G}|^{-1}\sum_{(i,t)\in\mathcal{G}} M_{it}\right)\left(\frac{\phi_{\breve{M},p,\sigma}(Y,\Omega)}{\phi_{M,p,\sigma}(Y,\Omega)} - 1\right)\right]$$
$$\leq \sqrt{\mathbb{E}_{M,p,\sigma}\left[U(Y,\Omega) - |\mathcal{G}|^{-1}\sum_{(i,t)\in\mathcal{G}} M_{it}\right]^2} \times \sqrt{\mathbb{E}_{M,p,\sigma}\left[\left(\frac{\phi_{\breve{M},p,\sigma}(Y,\Omega)}{\phi_{M,p,\sigma}(Y,\Omega)} - 1\right)^2\right]}$$
$$= \sqrt{\mathbb{E}_{M,p,\sigma}\left[U(Y,\Omega) - |\mathcal{G}|^{-1}\sum_{(i,t)\in\mathcal{G}} M_{it}\right]^2} \times \sqrt{\mathbb{E}_{M,p,\sigma}\left[\left(\frac{\phi_{\breve{M},p,\sigma}(Y,\Omega)}{\phi_{M,p,\sigma}(Y,\Omega)}\right)^2\right] - 1}$$

where $\phi_{M,p,\sigma}(Y,\Omega)$ and $\phi_{\breve{M},p,\sigma}(Y,\Omega)$ are density functions of the observed data $\Omega \circ Y$ under (M,p,σ) and $(\breve{M},p,\sigma)$, respectively. It follows that

$$\mathbb{E}_{M,p,\sigma}\left[U(Y,\Omega) - |\mathcal{G}|^{-1}\sum_{(i,t)\in\mathcal{G}} M_{it}\right]^2 \geq \frac{\left(\alpha\sqrt{\chi_1+\chi_2} + o(s_*(M,p,\sigma))\right)^2}{\mathbb{E}_{M,p,\sigma}\left[\left(\frac{\phi_{\breve{M},p,\sigma}(Y,\Omega)}{\phi_{M,p,\sigma}(Y,\Omega)}\right)^2\right] - 1}.$$

Step 1: compute $\mathbb{E}_{M,p,\sigma}\left[\left(\frac{\phi_{\breve{M},p,\sigma}(Y,\Omega)}{\phi_{M,p,\sigma}(Y,\Omega)}\right)^2\right]$. Under the assumptions on Ω and $\mathcal{E}$, we have the following explicit formula for the density function:

$$\phi_{M,p,\sigma}(Y,\Omega) = (2\pi)^{-\frac{\sum_{i,t}\omega_{it}}{2}}\sigma^{-\sum_{i,t}\omega_{it}}\exp\left(-\frac{1}{2\sigma^2}\left\|Y-\Omega\circ M\right\|_F^2\right)p^{\sum_{i,t}\omega_{it}}(1-p)^{NT-\sum_{i,t}\omega_{it}}.$$

It implies that

$$\mathbb{E}_{M,p,\sigma}\left[\left(\frac{\phi_{\breve{M},p,\sigma}(Y,\Omega)}{\phi_{M,p,\sigma}(Y,\Omega)}\right)^2\right] = \mathbb{E}_{M,p,\sigma}\left[\exp\left(-\frac{1}{\sigma^2}\left\|Y-\Omega\circ\breve{M}\right\|_F^2 + \frac{1}{\sigma^2}\left\|Y-\Omega\circ M\right\|_F^2\right)\right]$$
$$= \mathbb{E}_{M,p,\sigma}\left[\exp\left(\sigma^{-2}\sum_{i\leq N,t\leq T}\omega_{it}\Delta_{it}^2\right)\right] = \exp\left(\sigma^{-2}\sum_{i\leq N,t\leq T}p\Delta_{it}^2\right)\mathbb{E}_{M,p,\sigma}\left[\exp\left(\sigma^{-2}\sum_{i\leq N,t\leq T}(\omega_{it}-p)\Delta_{it}^2\right)\right]$$

where $\Delta = \breve{M} - M = \breve{M}^\star - M^\star$ and the second equality holds under the assumptions on Ω and $\mathcal{E}$. In addition, following the same argument as Step 1 of the proof of Theorem 4.2 in Chernozhukov et al. (2021), we can show that $\sigma^{-2}\sum_{i\leq N,t\leq T}(\omega_{it} - p)\Delta_{it}^2$ has sub-Gaussian norm bounded by $C\sigma^{-2}\sqrt{\sum_{i\leq N,t\leq T}\Delta_{it}^4}$ for some constant $C>0$. Further, we have

$$\sum_{i\leq N,t\leq T}\Delta_{it}^4 \lesssim (N|\mathcal{T}|_o + T|\mathcal{I}|_o)\max_{it}\Delta_{it}^4$$
$$= (N|\mathcal{T}|_o + T|\mathcal{I}|_o)\max_{it}\left(\frac{\alpha}{\sqrt{\chi_1+\chi_2}}\left(\frac{\bar{F}_{\mathcal{T}}'(F'F)^{-1}F_t}{|\mathcal{I}|_o} + \frac{\beta_i'(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}}}{|\mathcal{T}|_o}\right) + \frac{\alpha^2}{\chi_1+\chi_2}\frac{\bar{F}_{\mathcal{T}}'(F'F)^{-1}}{|\mathcal{I}|_o}\frac{(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}}}{|\mathcal{T}|_o}\right)^4$$
$$\lesssim (N|\mathcal{T}|_o + T|\mathcal{I}|_o)(\alpha\sqrt{\chi_1+\chi_2} + o(\sqrt{\chi_1+\chi_2}))^4 \lesssim (N|\mathcal{T}|_o + T|\mathcal{I}|_o)\alpha^4(\chi_1+\chi_2)^2 \ll 1$$

where the last relation follows from $N|\mathcal{T}|_o \ll T^2|\mathcal{I}|_o^2$ and $T|\mathcal{I}|_o \ll N^2|\mathcal{T}|_o^2$. Then, Lemma B.3 of Chernozhukov et al. (2021) implies $\mathbb{E}_{M,p,\sigma}\left[\exp\left(\sigma^{-2}\sum_{i\leq N,t\leq T}(\omega_{it}-p)\Delta_{it}^2\right)\right] = 1 + o(1)$, and Claim F.9 (ii) yields $\exp\left(\sigma^{-2}\sum_{i\leq N,t\leq T}p\Delta_{it}^2\right) = \exp\left(\sigma^{-2}p\left\|\Delta\right\|_F^2\right) = \exp\left(\sigma^{-2}p\alpha^2(1+o(1))\right)$. Therefore, we reach $\mathbb{E}_{M,p,\sigma}\left[\left(\frac{\phi_{\breve{M},p,\sigma}(Y,\Omega)}{\phi_{M,p,\sigma}(Y,\Omega)}\right)^2\right] = \exp\left(\sigma^{-2}p\alpha^2(1+o(1))\right)(1+$

$o(1))$.

Step 2: derive the final result. Since α is a fixed constant and $s_*(M,p,\sigma) \asymp \sqrt{\chi_1+\chi_2}$, we have $(\alpha\sqrt{\chi_1+\chi_2}+o(s_*(M,p,\sigma)))^2 = \alpha^2(\chi_1+\chi_2)(1+o(1))$. In addition, since $\exp(a)-1 \le a+\exp(a)a^2/2$ for any $a>0$, we have

$$\exp\left(\sigma^{-2}p\alpha^2(1+o(1))\right)-1 \le \sigma^{-2}p\alpha^2(1+o(1))+\exp\left(\sigma^{-2}p\alpha^2(1+o(1))\right)(\sigma^{-2}p\alpha^2(1+o(1)))^2/2.$$

Therefore, for $\alpha \in (0, \frac{\sigma}{4\sqrt{p}})$, $\exp\left(\sigma^{-2}p\alpha^2(1+o(1))\right)-1 \le \sigma^{-2}p\alpha^2(1+o(1))+\exp(1/4)O(\sigma^{-4}p^2\alpha^4)$. Using these bounds, we reach

$$\begin{aligned}\mathbb{E}_{M,p,\sigma}\left[U(Y,\Omega)-|\mathcal{G}|^{-1}\sum_{(i,t)\in\mathcal{G}}M_{it}\right]^2 &\ge \frac{(\chi_1+\chi_2)(1+o(1))}{\sigma^{-2}p(1+o(1))+\exp(1/4)O(\sigma^{-4}p^2\alpha^2)+\exp(1/4)o(1)}\\ &= \frac{(\chi_1+\chi_2)(1+o(1))}{\sigma^{-2}p(1+o(1))+\exp(1/4)O(\sigma^{-4}p^2\alpha^2)} = s_*^2(M,p,\sigma)\frac{1+o(1)}{1+o(1)+O(\sigma^{-2}p\alpha^2)}\end{aligned}$$

where the third line holds since $\sigma^{-2}p$ is assumed to be bounded. As a result, we have

$$\frac{\mathbb{E}_{M,p,\sigma}\left[U(Y,\Omega)-|\mathcal{G}|^{-1}\sum_{(i,t)\in\mathcal{G}}M_{it}\right]^2}{s_*^2(M,p,\sigma)} \ge \frac{1+o(1)}{1+o(1)+O(\sigma^{-2}p\alpha^2)}$$

for any $\alpha \in (0, \frac{\sigma}{4\sqrt{p}})$. By taking sufficiently small $\alpha>0$, we attain

$$\liminf_{N,T\to\infty}\frac{\mathbb{E}_{M,p,\sigma}\left[U(Y,\Omega)-|\mathcal{G}|^{-1}\sum_{(i,t)\in\mathcal{G}}M_{it}\right]^2}{s_*^2(M,p,\sigma)} \ge 1.$$

F.3 Proofs for Section 4

F.3.1 Proof of Theorem 4.1

For each potential realization $\iota \in \{0,1\}$, Assumptions Assumptions 3.2, 3.3 and C.1-C.3 are satisfied. Hence, We can use the results of Propositions E.1, F.1 and Theorem C.1 for each potential realization $\iota \in \{0,1\}$. Then, following the proof of Theorem C.1, for each $\iota \in \{0,1\}$, we have $\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\left(\widehat{\beta}_i^{(\iota)\prime}\widehat{F}_t^{(\iota)}-M_{it}^{(\iota)}\right)=P_1^{(\iota)}+P_2^{(\iota)}$ where

$$\begin{aligned}P_1^{(\iota)} &:= \frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\left(\beta_i'\left(\sum_{j=1}^N\omega_{jt}^{(\iota)}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N\omega_{jt}^{(\iota)}\beta_j\varepsilon_{jt}\right)+F_t^{(\iota)\prime}\left(\sum_{s=1}^T\omega_{is}^{(\iota)}F_s^{(\iota)}F_s^{(\iota)\prime}\right)^{-1}\left(\sum_{s=1}^T\omega_{is}^{(\iota)}F_s^{(\iota)}\varepsilon_{is}\right)\right),\\ P_2^{(\iota)} &:= \left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{\frac{1}{2}}\sum_{d=1}^7\mathcal{R}_{d,\mathcal{G}}^{(\iota)},\\ \mathcal{V}_{\mathcal{G}}^{(\iota)} &= \frac{1}{|\mathcal{T}|_o^2}\sum_{t\in\mathcal{T}}\bar{\beta}_{\mathcal{I}}'\left(\sum_{j=1}^N\omega_{jt}^{(\iota)}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N\omega_{jt}^{(\iota)}\sigma_{jt}^2\beta_j\beta_j'\right)\left(\sum_{j=1}^N\omega_{jt}^{(\iota)}\beta_j\beta_j'\right)^{-1}\bar{\beta}_{\mathcal{I}}\\ &\quad+\frac{1}{|\mathcal{I}|_o^2}\sum_{i\in\mathcal{I}}\bar{F}_{\mathcal{T}}^{(\iota)\prime}\left(\sum_{s=1}^T\omega_{is}^{(\iota)}F_s^{(\iota)}F_s^{(\iota)\prime}\right)^{-1}\left(\sum_{s=1}^T\omega_{is}^{(\iota)}\sigma_{is}^2F_s^{(\iota)}F_s^{(\iota)\prime}\right)\left(\sum_{s=1}^T\omega_{is}^{(\iota)}F_s^{(\iota)}F_s^{(\iota)\prime}\right)^{-1}\bar{F}_{\mathcal{T}}^{(\iota)},\end{aligned}$$

and $\sum_{d=1}^7\left|\mathcal{R}_{d,\mathcal{G}}^{(\iota)}\right|=o_P(1)$. Then, we have

$$\left(\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}\left(\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\widehat{\Gamma}_{it}-\Gamma_{it}\right)=\left(\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}\left(P_1^{(1)}-P_1^{(0)}\right)+\left(\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}\left(P_2^{(1)}-P_2^{(0)}\right).$$

We first derive $\left(\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}\left(P_1^{(1)}-P_1^{(0)}\right)\xrightarrow{D}\mathcal{N}(0,1)$. Let $\Phi_{NT}^{(0)}\coloneqq\left(\mathcal{V}_{\mathcal{G}}^{(0)}\right)^{-\frac{1}{2}}P_1^{(0)}$ and $\Phi_{NT}^{(1)}\coloneqq\left(\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}P_1^{(1)}$. Then, by Step 3 of the proof of Theorem C.1, we know that $\Phi_{NT}^{(0)}\xrightarrow{D}\mathcal{N}(0,1)$ and $\Phi_{NT}^{(1)}\xrightarrow{D}\mathcal{N}(0,1)$. Furthermore, we can show that $\left(\Phi_{NT}^{(0)},\Phi_{NT}^{(1)}\right)$ converges (jointly) to a bivariate normal distribution. Note that, by using the same method in Step 3 of the proof of Theorem C.1, for each $\iota\in\{0,1\}$,

$$\left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\beta_i'\left(\sum_{j=1}^N\omega_{jt}^{(\iota)}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N\omega_{jt}^{(\iota)}\beta_j\varepsilon_{jt}\right)=\left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{-\frac{1}{2}}\sum_{s=1}^T\sum_{j=1}^N\frac{\varepsilon_{js}\omega_{js}^{(\iota)}}{|\mathcal{T}|_o}\bar{\beta}_{\mathcal{I}}'\left(\Xi_{1,s}^{(\iota)}\right)^{-1}\beta_j1\{s\in\mathcal{T}\},$$

$$\left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}F_t^{(\iota)\prime}\left(\sum_{s=1}^T\omega_{is}^{(\iota)}F_s^{(\iota)}F_s^{(\iota)\prime}\right)^{-1}\left(\sum_{s=1}^T\omega_{is}^{(\iota)}F_s^{(\iota)}\varepsilon_{is}\right)=\left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{-\frac{1}{2}}\sum_{j=1}^N\sum_{s=1}^T\frac{\varepsilon_{js}\omega_{js}^{(\iota)}}{|\mathcal{I}|_o}\bar{F}_{\mathcal{T}}^{(\iota)\prime}\left(\Xi_{2,j}^{(\iota)}\right)^{-1}F_s^{(\iota)}1\{j\in\mathcal{I}\},$$

where $\Xi_{1,t}^{(\iota)}\coloneqq\sum_{j=1}^N\omega_{jt}^{(\iota)}\beta_j\beta_j'$ and $\Xi_{2,i}^{(\iota)}\coloneqq\sum_{s=1}^T\omega_{is}^{(\iota)}F_s^{(\iota)}F_s^{(\iota)\prime}$. Hence, for each $\iota\in\{0,1\}$, we have $\Phi_{NT}^{(\iota)}=\sum_{j=1}^N\sum_{s=1}^T\varepsilon_{js}\omega_{js}^{(\iota)}A_{js}^{(\iota)}\overset{(a)}{=}\sum_{j=1}^N\sum_{s=1}^T\varepsilon_{js}\omega_{js}^{(\iota)}A_{js}^{(\iota)}1\{(j,s)\in\chi\}$ where $A_{js}^{(\iota)}=\left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{-\frac{1}{2}}\frac{1}{|\mathcal{T}|_o}\bar{\beta}_{\mathcal{I}}'\left(\Xi_{1,s}^{(\iota)}\right)^{-1}\beta_j1\{s\in\mathcal{T}\}+\left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{-\frac{1}{2}}\frac{1}{|\mathcal{I}|_o}\bar{F}_{\mathcal{T}}^{(\iota)\prime}\left(\Xi_{2,j}^{(\iota)}\right)^{-1}F_s^{(\iota)}1\{j\in\mathcal{I}\}$ and $\chi=\{(i,t):t\in\mathcal{T}\text{ or }i\in\mathcal{I}\}$. Here, Equation (a) comes from the fact that $A_{js}^{(\iota)}\neq0$ only when $s\in\mathcal{T}$ or $j\in\mathcal{I}$. Then, for any constants a, b, we have $a\Phi_{NT}^{(0)}+b\Phi_{NT}^{(1)}=\sum_{j=1}^N\sum_{s=1}^Ta\varepsilon_{js}\omega_{js}^{(0)}A_{js}^{(0)}1\{(j,s)\in\chi\}+b\varepsilon_{js}\omega_{js}^{(1)}A_{js}^{(1)}1\{(j,s)\in\chi\}$. Some calculation shows that $\text{Var}\left(a\Phi_{NT}^{(0)}+b\Phi_{NT}^{(1)}\middle|\zeta,\Upsilon\right)\xrightarrow{P}a^2+b^2$.

Next, we check the Lindeberg's condition. Let $\mathcal{Y}_{js}\coloneqq a\varepsilon_{js}\omega_{js}^{(0)}A_{js}^{(0)}1\{(j,s)\in\chi\}+b\varepsilon_{js}\omega_{js}^{(1)}A_{js}^{(1)}1\{(j,s)\in\chi\}$, so that $a\Phi_{NT}^{(0)}+b\Phi_{NT}^{(1)}=\sum_{j=1}^N\sum_{s=1}^T\mathcal{Y}_{js}$. Then, we have

$$\begin{aligned}\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\mathcal{Y}_{js}^4\middle|\zeta,\Upsilon\right]&=\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\left(a\varepsilon_{js}\omega_{js}^{(0)}A_{js}^{(0)}1\{(j,s)\in\chi\}+b\varepsilon_{js}\omega_{js}^{(1)}A_{js}^{(1)}1\{(j,s)\in\chi\}\right)^4\middle|\zeta,\Upsilon\right]\\&=\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\left(a\varepsilon_{js}\omega_{js}^{(0)}A_{js}^{(0)}1\{(j,s)\in\chi\}\right)^4\middle|\zeta,\Upsilon\right]+\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\left(b\varepsilon_{js}\omega_{js}^{(1)}A_{js}^{(1)}1\{(j,s)\in\chi\}\right)^4\middle|\zeta,\Upsilon\right],\end{aligned}$$

where the last equality comes from the fact that the cross-product terms are zero because $\omega_{js}^{(0)}\omega_{js}^{(1)}=(1-\Upsilon_{js})\Upsilon_{js}=0$. Then, we have for $\iota\in\{0,1\}$,

$$\begin{aligned}&\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\left(\varepsilon_{js}\omega_{js}^{(\iota)}A_{js}^{(\iota)}1\{(j,s)\in\chi\}\right)^4\middle|\zeta,\Upsilon\right]=\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\varepsilon_{js}^4\middle|\zeta,\Upsilon\right]\omega_{js}^{(\iota)}\left(A_{js}^{(\iota)}\right)^41\{(j,s)\in\chi\}\\&\le C\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\varepsilon_{js}^4\middle|\zeta,\Upsilon\right]\left(\mathcal{V}_{\mathcal{G}}^{(\iota)}\right)^{-2}\omega_{js}^{(\iota)}\left(\frac{1}{|\mathcal{T}|_o^4}\left(\bar{\beta}_{\mathcal{I}}'\left(\Xi_{1,s}^{(\iota)}\right)^{-1}\beta_j\right)^41\{s\in\mathcal{T}\}+\frac{1}{|\mathcal{I}|_o^4}\left(\bar{F}_{\mathcal{T}}^{(\iota)\prime}\left(\Xi_{2,j}^{(\iota)}\right)^{-1}F_s^{(\iota)}\right)^41\{j\in\mathcal{I}\}\right)\\&=O_P\left(\frac{p_{\max}^{(\iota)2}\varpi^4K^4}{p_{\min}^{(\iota)4}N|\mathcal{T}|_o}+\frac{p_{\max}^{(\iota)2}\left(q^{(\iota)}\right)^8\varpi^4K^4}{p_{\min}^{(\iota)4}T|\mathcal{I}|_o}\right).\end{aligned}$$

Using the above bound, we can have the following Lindeberg's condition:

$$\text{Var}\left(a\Phi_{NT}^{(0)}+b\Phi_{NT}^{(1)}|\zeta,\Upsilon\right)^{-1}\sum_{j=1}^N\sum_{s=1}^T\mathbb{E}\left[\mathcal{Y}_{js}^21\left\{|\mathcal{Y}_{jt}|>\epsilon\text{Var}\left(a\Phi_{NT}^{(0)}+b\Phi_{NT}^{(1)}|\zeta,\Upsilon\right)^{\frac{1}{2}}\right\}\middle|\zeta,\Upsilon\right]=o_P(1)$$

Hence, by the central limit theorem, we have, conditioning on $\{\zeta,\Upsilon\}$, $a\Phi_{NT}^{(0)}+b\Phi_{NT}^{(1)}\xrightarrow{D}a\Phi^{(0)}+b\Phi^{(1)}$ where the two "Φ" terms are jointly standard normal. From here, by using the same argument as in the proof of Theorem 3 in Bai (2003), we have conditioning on $\{\zeta, \Upsilon\}$,

$$\left(\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}\left(P_1^{(1)}-P_1^{(0)}\right)=\frac{\left(\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{\frac{1}{2}}\Phi_{NT}^{(1)}+\left(\mathcal{V}_{\mathcal{G}}^{(0)}\right)^{\frac{1}{2}}\Phi_{NT}^{(0)}}{\sqrt{\mathcal{V}_{\mathcal{G}}^{(1)}\mathrm{Var}\left(\Phi^{(1)}|\zeta,\Upsilon\right)+\mathcal{V}_{\mathcal{G}}^{(0)}\mathrm{Var}\left(\Phi^{(0)}|\zeta,\Upsilon\right)}}\xrightarrow{D}\mathcal{N}(0,1).$$

Therefore, we have $\left(\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}\left(P_1^{(1)}-P_1^{(0)}\right)+\left(\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}\right)^{-\frac{1}{2}}\left(P_2^{(1)}-P_2^{(0)}\right)\xrightarrow{D}\mathcal{N}(0,1)$. □

F.3.2 Proof of Corollary 4.2

It is straightforward to prove $\frac{\mathcal{V}_{\mathcal{G}}^{(0)}+\mathcal{V}_{\mathcal{G}}^{(1)}}{\widehat{\mathcal{V}}_{\mathcal{G}}^{(0)}+\widehat{\mathcal{V}}_{\mathcal{G}}^{(1)}}\xrightarrow{P}1$. It then leads to the result.

F.4 Technical Claims for Section F

Claim F.1. *(i) There is a $K\times K$ matrix H_1 such that $\frac{1}{\sqrt{N}}\beta H_1$ is the left singular vector of $M^\star$.*
(ii) Let $\widetilde{\psi}_r$, ψ_r be the rth largest singular value of $\widetilde{M}$ and $M^\star$ respectively. Then, we have with probability converging to 1, $\min_{1\le r\le K}\psi_{\min}^{-2}\left|\psi_{r-1}^2-\widetilde{\psi}_r^2\right|\ge\bar{c}/2$, $\min_{1\le r\le K}\psi_{\min}^{-2}\left|\widetilde{\psi}_r^2-\psi_{r+1}^2\right|\ge\bar{c}/2$, where $\bar{c}=\min_{1\le r\le K+1}\left|c_{r-1}^2-c_r^2\right|$ and $c_r=\psi_r/\psi_{\min}$.
(iii) $P\left(\widehat{K}=K\right)\to 1$. (iv) $\|\frac{1}{\sqrt{N}}\widetilde{\beta}-\frac{1}{\sqrt{N}}\beta H_1\|_F=O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}}c_{\mathrm{inv}}q^2K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}}\right)$.

Proof of Claim F.1. (i) Let $L=(\beta'\beta)^{\frac{1}{2}}F'F(\beta'\beta)^{\frac{1}{2}}$. Let G_L be a $K\times K$ matrix whose columns are the eigenvectors of L such that $\Lambda_L=G_L'LG_L$ is a descending order diagonal matrix of the eigenvalues of L. Define $H_1:=\sqrt{N}(\beta'\beta)^{-\frac{1}{2}}G_L$. Then, we have $\frac{1}{\sqrt{N}}(\beta F'F\beta')\beta H_1=\frac{1}{\sqrt{N}}\beta(\beta'\beta)^{-\frac{1}{2}}(\beta'\beta)^{\frac{1}{2}}F'F(\beta'\beta)^{\frac{1}{2}}(\beta'\beta)^{\frac{1}{2}}H_1=\beta(\beta'\beta)^{-\frac{1}{2}}LG_L=\frac{1}{\sqrt{N}}\beta H_1\Lambda_L$. Note that $\frac{1}{N}(\beta H_1)'\beta H_1=\frac{1}{N}H_1'\beta'\beta H_1=G_L'(\beta'\beta)^{-\frac{1}{2}}\beta'\beta(\beta'\beta)^{-\frac{1}{2}}G_L=G_L'G_L=I_K$. So the column of $\frac{1}{\sqrt{N}}\beta H_1$ are the eigenvector of $\beta F'F\beta'$ corresponding to the eigenvalues of Λ_L. Hence, $\frac{1}{\sqrt{N}}\beta H_1$ is the left singular vector of $M^\star=\beta F'$ (Note that a singular vector of a nonzero singular value is unique up to sign. By changing the sign of each column of G_L, we can change the sign of each column of $\frac{1}{\sqrt{N}}\beta H_1$ to correspond with the sign of each column of $\frac{1}{\sqrt{N}}\widetilde{\beta}$).

Parts (ii) (iii) (iv) follow from a straightforward application of Weyl's theorem and Davis-Kahan theorem (ex. Yu et al. (2015)). So we omit the details for brevity. □

Claim F.2. *(i) There is a $K\times K$ matrix $H_4^{[t+N]}$ such that $\psi_{\min}^{-1/2}\widetilde{W}^{[t+N]}H_4^{[t+N]}$ is the left singular vector of $\widetilde{W}^{[t+N]}\widetilde{Z}^{[t+N]\prime}$.*
(ii) Let $\widetilde{\psi}_r$, $\widetilde{\phi}_r^{[t]}$ be the rth largest singular value of $\widetilde{M}$ and $\widetilde{W}^{[t+N]}\widetilde{Z}^{[t+N]\prime}$ respectively. Then, we have with probability converging to 1, $\min_{1\le t\le T}\min_{1\le r\le K}\psi_{\min}^{-2}\left|\widetilde{\psi}_{r-1}^2-\widetilde{\phi}_r^{[t]2}\right|\ge\bar{c}/2$, $\min_{1\le t\le T}\min_{1\le r\le K}\psi_{\min}^{-2}\left|\widetilde{\phi}_r^{[t]2}-\widetilde{\psi}_{r+1}^2\right|\ge\bar{c}/2$, where $\bar{c}=\min_{1\le r\le K+1}\left|c_{r-1}^2-c_r^2\right|$ and $c_r=\psi_r/\psi_{\min}$.
(iii) $\max_{1\le t\le T}\left\|\frac{1}{\sqrt{N}}\widetilde{\beta}-\psi_{\min}^{-1/2}\widetilde{W}^{[t+N]}H_4^{[t+N]}\right\|_F=O_P\left(\frac{\sigma p_{\max}\vartheta^{\frac{1}{2}}c_{\mathrm{inv}}q^{\frac{9}{2}}\mu^{\frac{1}{2}}K^{\frac{3}{2}}\max\{\sqrt{N\log N},\sqrt{T\log T}\}}{p_{\min}^2\min\{\sqrt{N},\sqrt{T}\}\psi_{\min}}\right)$.

Proof of Claim F.2. The proof is similar to that of Claim F.1. So, we omit it. □

Claim F.3. *(i) $\|\beta H_1\|=\sqrt{N}$, $\max_i\|H_1'\beta_i\|\le\mu^{\frac{1}{2}}K^{\frac{1}{2}}$, (ii) $\|FH_1'^{-1}\|=\frac{q\psi_{\min}}{\sqrt{N}}$, $\max_t\|H_1^{-1}F_t\|\le\frac{q\mu^{\frac{1}{2}}K^{\frac{1}{2}}\psi_{\min}}{\sqrt{NT}}$, (iii) $\|\beta H_2^{-1}\|=O_P\left(\sqrt{N}\right)$, $\max_i\|H_2'^{-1}\beta_i\|=O_P\left(\mu^{\frac{1}{2}}K^{\frac{1}{2}}\right)$, (iv) $\|FH_2'\|=O_P\left(\frac{q\psi_{\min}}{\sqrt{N}}\right)$, $\max_t\|H_2F_t\|=O_P\left(\frac{q\mu^{\frac{1}{2}}K^{\frac{1}{2}}\psi_{\min}}{\sqrt{NT}}\right)$.*

Proof of Claim F.3. The proof is straightforward so we omit details for brevity.

Claim F.4. *(i) $\|B\|\le p_{max}$, $\|B^{-1}\|\le p_{\min}^{-1}$. (ii) $\max_t\|\widetilde{B}_t^{-1}-B^{-1}\|=O_P\left(\frac{\sigma p_{\max}^{\frac{3}{2}}c_{\mathrm{inv}}q^2K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}^3\psi_{\min}}+\frac{p_{\max}^{\frac{1}{2}}\vartheta\mu^{\frac{1}{2}}K\sqrt{\log T}}{p_{\min}^2\sqrt{N}}\right)$.*
(iii) $\max_t\|\widetilde{B}_t^{-1}\|=O_P(p_{\min}^{-1})$. (iv) $\max_t\|\widetilde{B}_t^{-1}-B_t^{*-1}\|=O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}}c_{\mathrm{inv}}q^2K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}^3\psi_{\min}}\right)$ and $\max_t\|B_t^{*-1}\|=O_P(p_{\min}^{-1})$.

Proof of Claim F.4. (i) Note that $B = \left(\frac{1}{\sqrt{N}}\beta H_1\right)' \Pi \left(\frac{1}{\sqrt{N}}\beta H_1\right) = U_{M^\star}'\Pi U_{M^\star}$. Hence, $\|B\| \leq \|U_{M^\star}\|^2\|\Pi\| = p_{max}$. Let $\psi_r(A)$ and $\psi_{\min}(A)$ be the rth largest singular value and non-zero smallest singular value of matrix A respectively. Because $\psi_K^2\left(U_{M^\star}'\Pi^{\frac{1}{2}}\right) = \psi_K\left(U_{M^\star}'\Pi U_{M^\star}\right) = \psi_K\left(\Pi^{\frac{1}{2}}U_{M^\star}U_{M^\star}'\Pi^{\frac{1}{2}}\right)$, we have $\psi_{\min}\left(U_{M^\star}'\Pi U_{M^\star}\right) = \psi_{\min}\left(\Pi^{\frac{1}{2}}U_{M^\star}U_{M^\star}'\Pi^{\frac{1}{2}}\right) \overset{(i)}{\geq} \psi_{\min}^2\left(\Pi^{\frac{1}{2}}\right)\psi_{\min}\left(U_{M^\star}U_{M^\star}'\right) = \psi_{\min}^2\left(\Pi^{\frac{1}{2}}\right)\psi_{\min}\left(U_{M^\star}'U_{M^\star}\right) = \psi_{\min}^2\left(\Pi^{\frac{1}{2}}\right) = p_{\min}$ where (i) comes from the fact that $\psi_{\min}(AB) \geq \psi_{\min}(A)\psi_{\min}(B)$ if A is a full column rank matrix and the fact that $\psi_{\min}(A) = \psi_{\min}(A')$ for any matrix A. So, we have $\|B^{-1}\| \leq p_{\min}^{-1}$.

(ii) Note that

$$\max_t \|\widetilde{B}_t - B\| = O_P(1)\left(\|\frac{1}{N}\sum_{j=1}^{N} p_j\left(\widetilde{\beta}_j - H_1'\beta_j\right)\beta_j'H_1\| + \|\frac{1}{N}\sum_{j=1}^{N} p_j H_1'\beta_j\left(\widetilde{\beta}_j - H_1'\beta_j\right)'\|\right)$$
$$+ O_P(1)\max_t \|\frac{1}{N}\sum_{j=1}^{N}\left(\omega_{jt} - p_j\right)H_1'\beta_j\beta_j'H_1\|.$$

The first and second terms are bounded like $\frac{1}{\sqrt{N}}\|\widetilde{\beta} - \beta H_1\|\|\Pi\|\frac{1}{\sqrt{N}}\|\beta H_1\| = O_P\left(\frac{\sigma p_{\max}^{\frac{3}{2}} c_{\text{inv}} q^2 K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}}\right)$ by Claim F.1 (iv). Moreover, we have $\|\frac{1}{N}\sum_{j=1}^N\left(\omega_{jt} - p_j\right)H_1'\beta_j\beta_j'H_1\| \leq \vartheta \max_g \|\sum_{j\in G_g}\left(\omega_{jt} - p_j\right)U_{M^\star,j}U_{M^\star,j}'\|$ by Lemma E.5. Hence, by matrix Bernstein inequality, $\max_t \|\frac{1}{N}\sum_{j=1}^N\left(\omega_{jt} - p_j\right)H_1'\beta_j\beta_j'H_1\| = O_P\left(\frac{p_{\max}^{\frac{1}{2}}\vartheta\mu^{\frac{1}{2}}K\sqrt{\log T}}{\sqrt{N}}\right)$ and $\max_t \|\widetilde{B}_t^{-1} - B^{-1}\| \leq \|B^{-1}\|^2\max_t \|\widetilde{B}_t - B\| = O_P\left(\frac{\sigma p_{\max}^{\frac{3}{2}} c_{\text{inv}} q^2 K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}^3\psi_{\min}} + \frac{p_{\max}^{\frac{1}{2}}\vartheta\mu^{\frac{1}{2}}K\sqrt{\log T}}{p_{\min}^2\sqrt{N}}\right)$.

(iii) It is easily derived from $\max_t \|\widetilde{B}_t^{-1}\| \leq \|B^{-1}\| + \max_t \|\widetilde{B}_t^{-1} - B^{-1}\|$.

(iv) The proof is the same as that of (ii).

Claim F.5. *(i)* $\max_t \|\left(\widetilde{W}^{[t+N]\prime}\widetilde{W}^{[t+N]}\right)^{-1}\widetilde{W}^{[t+N]\prime}\| = O_P\left(\psi_{\min}^{-1/2}\right)$,
(ii) $\max_t \|U_{M^\star} - \psi_{\min}^{-1/2}\widetilde{W}^{[t+N]}H_4^{[t+N]}\|_F = O_P\left(\frac{\sigma p_{\min}^{\frac{1}{2}} c_{\text{inv}} q^2 K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}}\right)$.

Proof of Claim F.5. (i) Note that

$$\max_t \frac{1}{\psi_{\min}\left(\widetilde{W}^{[t+N]}\right)} = \max_t \frac{1}{\psi_{\min}\left(\widetilde{W}^{[t+N]}\widetilde{H}^{[t+N]}\right)} \leq \frac{1}{\psi_{\min}(W)\left(1 - \psi_{\min}^{-1}(W)\max_t \|\widetilde{W}^{[t+N]}\widetilde{H}^{[t+N]} - W\|\right)}$$

Hence, we know that $\max_t \frac{1}{\psi_{\min}\left(\widetilde{W}^{[t+N]}\right)} = O_P\left(\psi_{\min}^{-1/2}\right)$. Then, we have

$$\max_t \|\left(\widetilde{W}^{[t+N]\prime}\widetilde{W}^{[t+N]}\right)^{-1}\widetilde{W}^{[t+N]\prime}\| = \max_t \psi_1^{1/2}\left(\left(\widetilde{W}^{[t+N]\prime}\widetilde{W}^{[t+N]}\right)^{-1}\right) = \max_t \frac{1}{\psi_{\min}\left(\widetilde{W}^{[t+N]}\right)} = O_P\left(\psi_{\min}^{-1/2}\right).$$

(ii) Since $\frac{1}{\sqrt{N}}\beta H_1 = U_{M^\star}$, we have by Claims F.1 (iv) and F.2 (iii)

$$\max_t \|U_{M^\star} - \psi_{\min}^{-1/2}\widetilde{W}^{[t+N]}H_4^{[t+N]}\|_F \leq \|U_{M^\star} - \frac{1}{\sqrt{N}}\widetilde{\beta}\|_F + \max_t \left\|\frac{1}{\sqrt{N}}\widetilde{\beta} - \psi_{\min}^{-1/2}\widetilde{W}^{[t+N]}H_4^{[t+N]}\right\|_F$$
$$= O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} c_{\text{inv}} q^2 K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}} + \frac{\sigma p_{\max}\vartheta^{\frac{1}{2}} c_{\text{inv}} q^{\frac{9}{2}}\mu^{\frac{1}{2}}K^{\frac{3}{2}}\max\{\sqrt{N\log N},\sqrt{T\log T}\}}{p_{\min}^2\min\{\sqrt{N},\sqrt{T}\}\psi_{\min}}\right)$$
$$= O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} c_{\text{inv}} q^2 K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}}\right). \quad \square$$

Claim F.6. *(i)* $\max_t \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\| = O_P\left(\sigma K^{\frac{1}{2}}\sqrt{N\log T}\right)$ *and* $\max_t \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_2'^{-1}\beta_j\| = O_P\left(\sigma K^{\frac{1}{2}}\sqrt{N\log T}\right)$.

(ii) $\max_i \|\sum_{s=1}^T \omega_{is}\varepsilon_{is}H_1^{-1}F_s\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} qK^{\frac{1}{2}}\sqrt{\log N}\psi_{\min}}{\sqrt{N}}\right)$ *and* $\max_i \|\sum_{s=1}^T \omega_{is}\varepsilon_{is}H_2F_s\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} qK^{\frac{1}{2}}\sqrt{\log N}\psi_{\min}}{\sqrt{N}}\right)$.

(iii) $\frac{1}{T}\sum_{t=1}^T \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\| = O_P(\sigma p_{\max}^{\frac{1}{2}} K^{\frac{1}{2}}\sqrt{N})$, $\frac{1}{|\mathcal{T}|_o}\sum_{t\in\mathcal{T}} \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\| = O_P(\sigma p_{\max}^{\frac{1}{2}} K^{\frac{1}{2}}\sqrt{N})$.

(iv) $\frac{1}{T}\sum_{t=1}^T \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\|^2 = O_P\left(\sigma^2 p_{\max} N\mu K\right)$.

Proof of Claim F.6. (i) First, note that $\beta H_1 = \sqrt{N}U_{M^\star}$. Then, by the matrix Bernstein inequality (ex. Koltchinskii et al. (2011)) with conditioning on $\{\mathcal{M},\Omega\}$, we have $\max_t \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\| = O_P\left(\sigma K^{\frac{1}{2}}\sqrt{N\log T}\right)$ and $\max_t \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_2'^{-1}\beta_j\| = O_P\left(\sigma K^{\frac{1}{2}}\sqrt{N\log T}\right)$ because $\|(I_K+\varphi')^{-1}\| = O_P(1)$.

(ii) Note that $FH_1'^{-1} = \frac{1}{\sqrt{N}}V_{M^\star}D_{M^\star}$ and that $\mathbb{E}[\varepsilon_{is}|\mathcal{M},\Omega] = 0$ and conditioning on $\{\mathcal{M},\Omega\}$, $\{\varepsilon_{is}\}$ are independent across s. Hence, conditioning on $\{\mathcal{M},\Omega\}$, by the matrix Bernstein inequality (ex. Koltchinskii et al. (2011)), we have $\max_i \|\sum_{s=1}^T \omega_{is}\varepsilon_{is}H_1^{-1}F_s\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} qK^{\frac{1}{2}}\sqrt{\log N}\psi_{\min}}{\sqrt{N}}\right)$. We can find the bound for $\max_i \|\sum_{s=1}^T \omega_{is}\varepsilon_{is}H_2F_s\|$ in the similar way to the proof of (i).

(iii) Note that $\frac{1}{T}\sum_{t=1}^T \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\| = \sqrt{N}\|U_{M^\star}\|_F\frac{1}{T}\sum_{t=1}^T \left\|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}\left(\frac{U_{M^\star,j}}{\|U_{M^\star}\|_F}\right)\right\|$. Let $L_t = \sum_{j=1}^N \omega_{jt}\varepsilon_{jt}\bar{U}_{M^\star,j}$ where $\bar{U}_{M^\star,j} = \frac{U_{M^\star,j}}{\|U_{M^\star}\|_F}$. In addition, we have $\mathbb{E}[\|L_t\|^2|\mathcal{M},\Omega] = \sum_{k=1}^K \mathbb{E}\left[\left(\sum_{j=1}^N \varepsilon_{jt}\omega_{jt}\bar{U}_{M^\star,j,k}\right)^2\middle|\mathcal{M},\Omega\right] = \sigma^2\sum_{k=1}^K\sum_{j=1}^N \omega_{jt}\bar{U}_{M^\star,j,k}^2$ and $\mathbb{E}[\|L_t\|^2|\mathcal{M}] \le C\sigma^2 p_{\max}$ where $\bar{U}_{M^\star,j,k}$ is the 'k'-th element of $\bar{U}_{M^\star,j}$, since by definition, $\|\bar{U}_{M^\star}\|_F = 1$. Then, we have $\mathbb{E}[\|L_t\||\mathcal{M}] \le C^{\frac{1}{2}}\sigma p_{\max}^{\frac{1}{2}}$ by Jensen's inequality. So, we have $\frac{1}{T}\sum_{t=1}^T \|L_t\| = O_P\left(\mathbb{E}[\frac{1}{T}\sum_{t=1}^T \|L_t\||\mathcal{M}]\right) = O_P\left(\sigma p_{\max}^{\frac{1}{2}}\right)$. Hence, we have $\frac{1}{T}\sum_{t=1}^T \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\| = O_P(\sigma p_{\max}^{\frac{1}{2}}K^{\frac{1}{2}}\sqrt{N})$. We can bound $\frac{1}{|\mathcal{T}|_o}\sum_{t\in\mathcal{T}} \|\sum_{j=1}^N \omega_{jt}\varepsilon_{jt}H_1'\beta_j\|$ by the same way.

(iv) The proof follows from straightforward calculating the expectation, so we omit the detail for brevity.

Claim F.7. *Let* $A_i^* := \frac{1}{T}\sum_{s=1}^T p_iH_2F_sF_s'H_2'$. *Then, we have (i)* $\max_i \|A_i^*\| = O_P\left(\frac{q^2\psi_{\min}^2}{NT}\right)$, $\max_i \|A_i^{*-1}\| = O_P\left(\frac{NT}{p_{\min}\psi_{\min}^2}\right)$. *(ii)* $\max_i \|A_i - A_i^*\| = O_P\left(\frac{p_{\max}q^2\mu^{\frac{1}{2}}K\sqrt{\log N}\psi_{\min}^2}{NT\sqrt{T}}\right)$, $\max_i \|A_i^{-1} - A_i^{*-1}\| = O_P\left(\frac{p_{\max}q^2\mu^{\frac{1}{2}}KN\sqrt{T\log N}}{p_{\min}^2\psi_{\min}^2}\right)$. *(iii)* $\frac{1}{T}\sum_{s=1}^T \|\widehat{F}_s - H_2F_s\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}}K^{\frac{1}{2}}}{p_{\min}\sqrt{N}}\right)$ *and* $\frac{1}{T}\sum_{s=1}^T \|\widehat{F}_s - H_2F_s\|^2 = O_P\left(\frac{\sigma^2 p_{\max}\mu K}{p_{\min}^2 N}\right)$. *(iv)* $\max_i \|\widehat{A}_i - A_i\| = O_P\left(\frac{\sigma p_{\max}q\mu K\psi_{\min}}{p_{\min}N\sqrt{T}}\right)$, $\max_i \|\widehat{A}_i^{-1} - A_i^{-1}\| = O_P\left(\frac{\sigma p_{\max}q\mu KNT^{\frac{3}{2}}}{p_{\min}^3\psi_{\min}^3}\right)$. *(v)* $\max_i \|A_i^{-1}\| = O_P\left(\frac{NT}{p_{\min}\psi_{\min}^2}\right)$ *and* $\max_i \|\widehat{A}_i^{-1}\| = O_P\left(\frac{NT}{p_{\min}\psi_{\min}^2}\right)$. *(vi)* $\max_s \|\widehat{F}_s - H_2F_s\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}}K^{\frac{1}{2}}\sqrt{\log T}}{p_{\min}\sqrt{N}}\right)$.

Proof of Claim F.7. (i) First, we know

$$\max_i \|A_i^*\| \le \frac{p_{max}}{T}\|H_2F'FH_2'\| = \frac{p_{max}}{NT}\|(I_K+\varphi)D_{M^\star}V_{M^\star}'V_{M^\star}D_{M^\star}(I_K+\varphi')\| \le \frac{p_{max}}{NT}\|D_{M^\star}^2\|\|I_K+\varphi\|^2 = O_P\left(\frac{q^2\psi_{\min}^2}{NT}\right),$$

because $\|I_K+\varphi\| = O_P(1)$ as noted in the proof of Claim F.3. Moreover, we have

$$\max_i \|A_i^{*-1}\| \le \max_i \frac{NT}{p_i}\|(I_K+\varphi')^{-1}D_{M^\star}^{-2}(I_K+\varphi)^{-1}\| \le \frac{NT}{p_{\min}}\psi_{\min}^{-2}\|(I_K+\varphi')^{-1}\|^2 = O_P\left(\frac{NT}{p_{\min}\psi_{\min}^2}\right).$$

(ii) Note that $A_i - A_i^* = (I_K+\varphi)\left(\frac{1}{T}\sum_{s=1}^T(\omega_{is}-p_i)H_1^{-1}F_sF_s'H_1'^{-1}\right)(I_K+\varphi')$. By the matrix Bernstein inequality with Claim F.3 (ii) $\max_i \|\sum_{s=1}^T(\omega_{is}-p_i)H_1^{-1}F_sF_s'H_1'^{-1}\| = O_P\left(\frac{p_{\max}q^2\mu^{\frac{1}{2}}K\sqrt{\log N}\psi_{\min}^2}{N\sqrt{T}}\right)$. Hence, we know $\max_i \|A_i - A_i^*\| = O_P\left(\frac{p_{\max}q^2\mu^{\frac{1}{2}}K\sqrt{\log N}\psi_{\min}^2}{NT\sqrt{T}}\right)$. By using the same method in the proof of Claim F.4 (ii), we have $\max_i \|A_i^{-1} - A_i^{*-1}\| = O_P\left(\frac{p_{\max}q^2\mu^{\frac{1}{2}}KN\sqrt{T\log N}}{p_{\min}^2\psi_{\min}^2}\right)$.

(iii), (iv) Because $\widehat{A}_i - A_i = \frac{1}{T}\sum_{s=1}^T\left[\omega_{is}(\widehat{F}_s - H_2F_s)F_s'H_2' + \omega_{is}H_2F_s(\widehat{F}_s - H_2F_s)' + \omega_{is}(\widehat{F}_s - H_2F_s)(\widehat{F}_s - H_2F_s)'\right]$, we know that $\max_i \|\widehat{A}_i - A_i\| = O_P(1)\max_i \|\frac{1}{T}\sum_{s=1}^T \omega_{is}(\widehat{F}_s - H_2F_s)F_s'H_2'\|$. By Proposition E.1, we know $\frac{1}{T}\sum_{s=1}^T \|\widehat{F}_s - H_2F_s\| =$

$O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} K^{\frac{1}{2}}}{p_{\min}\sqrt{N}}\right)$ and $\frac{1}{T}\sum_{s=1}^T \|\widehat{F}_s - H_2F_s\|^2 = O_P\left(\frac{\sigma^2 p_{\max}\mu K}{p_{\min}^2 N}\right)$. In addition, we have

$$\max_i \|\widehat{A}_i - A_i\| = O_P(p_{\max}^{\frac{1}{2}}) \max_s \|F_s' H_2'\| \left(\frac{1}{T}\sum_{s=1}^T \|\widehat{F}_s - H_2F_s\|^2\right)^{\frac{1}{2}} = O_P\left(\frac{\sigma p_{\max} q\mu K\psi_{\min}}{p_{\max}N\sqrt{T}}\right).$$

By the same method above, we derive $\max_i \|\widehat{A}_i^{-1} - A_i^{-1}\| \leq O_P(1)\max_i \|A_i^{-1}\|^2 \max_i \|\widehat{A}_i - A_i\|$.
(v) $\max_i \|A_i^{-1}\| = O_P\left(\frac{NT}{p_{\min}\psi_{\min}^2}\right)$ is proved in the above. $\max_i \|\widehat{A}_i^{-1}\| = O_P\left(\frac{NT}{p_{\min}\psi_{\min}^2}\right)$ follows from the relation $\max_i \|\widehat{A}_i^{-1}\| \leq \max_i \|\widehat{A}_i^{-1} - A_i^{-1}\| + \max_i \|A_i^{-1}\|$.
(vi) It follows from Proposition E.1.

Claim F.8. *Under the assumption of Theorem 3.2, we have*
(i) $\max_i \|\widehat{\beta}_i - H_2'^{-1}\beta_i\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} qK^{\frac{1}{2}}\sqrt{N\log N}}{p_{\min}\psi_{\min}}\right)$. *(ii)* $\max_i |\widehat{\sigma}_i{}^2 - \sigma_i^2| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} q^2\mu^{\frac{1}{2}} K\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}\min\{\sqrt{N},\sqrt{T}\}}\right)$.
(iii) $\max_t \left\|\left(\sum_{j=1}^N \omega_{jt}\widehat{\beta}_j\widehat{\beta}_j'\right)^{-1} - \left(\sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right)^{-1}\right\| = O_P\left(\frac{\sigma p_{\max} qK^{\frac{1}{2}}\sqrt{\log N}}{p_{\min}^3\sqrt{N}\psi_{\min}}\right)$.
(iv) $\max_i \left\|\left(\sum_{s=1}^T \omega_{is}\widehat{F}_s\widehat{F}_s'\right)^{-1} - \left(\sum_{s=1}^T \omega_{is}H_2F_sF_s'H_2'\right)^{-1}\right\| = O_P\left(\frac{\sigma p_{\max} qK^{\frac{1}{2}}N\sqrt{T\log T}}{p_{\min}^3\psi_{\min}^3}\right)$.

Proof of Claim F.8. (i) By Proposition F.1, Claims F.6 and F.7, we have

$$\max_i \|\widehat{\beta}_i - H_2'^{-1}\beta_i\| \leq \max_i \|H_2'^{-1}\left(\sum_{s=1}^T \omega_{is}F_sF_s'\right)^{-1} H_2^{-1}\| \max_i \|\sum_{s=1}^T \omega_{is}\varepsilon_{is}H_2F_s\| + \max_i \|R_i^{\beta}\| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} qK^{\frac{1}{2}}\sqrt{N\log N}}{p_{\min}\psi_{\min}}\right).$$

(ii) Because $\widehat{\sigma}_j^2 = \frac{1}{|\mathcal{W}_j|_o}\sum_{t\in\mathcal{W}_j}\widehat{\varepsilon}_{jt}^2$ where $\mathcal{W}_j = \{t : \omega_{jt} = 1\}$ and $\widehat{\varepsilon}_{it} = y_{it} - \widehat{\beta}_i'\widehat{F}_t$, we have $|\widehat{\sigma}_j^2 - \sigma_j^2| \leq \max_j |\frac{1}{|\mathcal{W}_j|_o}\sum_{t\in\mathcal{W}_j}\widehat{\varepsilon}_{jt}^2 - \varepsilon_{jt}^2| + \max_j |\frac{1}{|\mathcal{W}_j|_o}\sum_{t\in\mathcal{W}_j}\varepsilon_{jt}^2 - \sigma_j^2|$. We bound the first term. We have

$$\max_j \left|\frac{1}{|\mathcal{W}_j|_o}\sum_{t\in\mathcal{W}_j}\widehat{\varepsilon}_{jt}^2 - \varepsilon_{jt}^2\right| \leq 2\max_{i,t}|M_{it}^R|\max_j \frac{1}{|\mathcal{W}_j|_o}\sum_{t\in\mathcal{W}_j}|\varepsilon_{jt}| + \max_{i,t}\left(M_{it}^R\right)^2 + \max_{i,t}\left|\beta_i'F_t - \widehat{\beta}_i'\widehat{F}_t\right|^2$$
$$+ 2\max_{i,t}\left|\beta_i'F_t - \widehat{\beta}_i'\widehat{F}_t\right|\max_j \frac{1}{|\mathcal{W}_j|_o}\sum_{t\in\mathcal{W}_j}|\varepsilon_{jt}| + 2\max_{i,t}|M_{it}^R|\max_{i,t}\left|\beta_i'F_t - \widehat{\beta}_i'\widehat{F}_t\right| = O_P(\sigma)\left(\max_{i,t}|M_{it}^R| + \max_{i,t}\left|\beta_i'F_t - \widehat{\beta}_i'\widehat{F}_t\right|\right).$$

From Claims F.3, F.7 and F.8, we have $\max_{i,t}\left|\beta_i'F_t - \widehat{\beta}_i'\widehat{F}_t\right| = O_P\left(\frac{\sigma p_{\max}^{\frac{1}{2}} q^2\mu^{\frac{1}{2}} K\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}\min\{\sqrt{N},\sqrt{T}\}}\right)$. This implies the first term is $O_P\left(\frac{\sigma^2 p_{\max}^{\frac{1}{2}} q^2\mu^{\frac{1}{2}} K\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}\min\{\sqrt{N},\sqrt{T}\}}\right)$. In addition, since $\sum_{t\in\mathcal{W}_j}\varepsilon_{jt}^2 - \sigma_j^2 = \sum_{t\leq T}\omega_{jt}\varepsilon_{jt}^2 - \omega_{jt}\sigma_j^2$ and $|\mathcal{W}_j|_o = \sum_t \omega_{jt}$, the second term is $O_p\left(\frac{\sigma^2 p_{\max}^{1/2}\sqrt{\log T}}{p_{\min}\sqrt{T}}\right)$ by the matrix Bernstein inequality.
(iii) Because $\max_t \left\|\sum_{j=1}^N \omega_{jt}\left(\widehat{\beta}_j - H_2'^{-1}\beta_j\right)\beta_j'H_2^{-1}\right\| \leq \max_j \left\|\widehat{\beta}_j - H_2'^{-1}\beta_j\right\|\sqrt{N}p_{\max}^{\frac{1}{2}}\left\|\beta H_2^{-1}\right\| = O_P\left(\frac{\sigma p_{\max} qK^{\frac{1}{2}}N^{\frac{3}{2}}\sqrt{\log N}}{p_{\min}\psi_{\min}}\right)$ by Claims F.3 and F.8, we have

$$\max_t \left\|\sum_{j=1}^N \omega_{jt}\widehat{\beta}_j\widehat{\beta}_j' - \sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right\| = O_P(1)\max_t \left\|\sum_{j=1}^N \omega_{jt}\left(\widehat{\beta}_j - H_2'^{-1}\beta_j\right)\beta_j'H_2^{-1}\right\| = O_P\left(\frac{\sigma p_{\max} qK^{\frac{1}{2}}N^{\frac{3}{2}}\sqrt{\log N}}{p_{\min}\psi_{\min}}\right).$$

Then, by using the same assertion in the proof of Claim F.4 (ii), we have

$$\max_t \left\|\left(\sum_{j=1}^N \omega_{jt}\widehat{\beta}_j\widehat{\beta}_j'\right)^{-1} - \left(\sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right)^{-1}\right\|$$

$$= O_P(1)\max_t \left\|\sum_{j=1}^N \omega_{jt}\widehat{\beta}_j\widehat{\beta}_j' - \sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right\| \max_t \left\|\left(\sum_{j=1}^N \omega_{jt}H_2'^{-1}\beta_j\beta_j'H_2^{-1}\right)^{-1}\right\|^2 = O_P\left(\frac{\sigma p_{\max} q K^{\frac{1}{2}}\sqrt{\log N}}{p_{\min}^3\sqrt{N}\psi_{\min}}\right).$$

(iv) Using the same method of the proof of (iii), we can prove it. We omit the proof to save space. □

Claim F.9. *Under the assumptions of Theorem 3.3 and* $\max\{\psi_{\min}^{-1}(S_\beta), \psi_{\min}^{-1}(S_F)\} < \eta$*, we have*
(i) $s_*^2(\breve{M}, p, \sigma) = s_*^2(M, p, \sigma) + o(s_*^2(M, p, \sigma))$*, and*
(ii) $\|\Delta\|_F^2 = \alpha^2(1 + o(1))$.

Proof of Claim F.9. (i) To begin with, we observe that

$$\bar{\breve{\beta}}_{\mathcal{I}} - \bar{\beta}_{\mathcal{I}} = \frac{\alpha}{\sqrt{\chi_1+\chi_2}}\frac{(F'F)^{-1}\bar{F}_{\mathcal{T}}}{|\mathcal{I}|_o} = O(\frac{\alpha}{|\mathcal{I}|_o T}\frac{\eta}{\sqrt{\chi_1+\chi_2}}),$$

and

$$\breve{\beta}'\breve{\beta} - \beta'\beta = \frac{\alpha(F'F)^{-1}\bar{F}_{\mathcal{T}}\bar{\beta}_{\mathcal{I}}'}{\sqrt{\chi_1+\chi_2}} + \frac{\alpha\bar{\beta}_{\mathcal{I}}\bar{F}_{\mathcal{T}}'(F'F)^{-1}}{\sqrt{\chi_1+\chi_2}} + \frac{\alpha^2(F'F)^{-1}\bar{F}_{\mathcal{T}}\bar{F}_{\mathcal{T}}'(F'F)^{-1}}{|\mathcal{I}|_o(\chi_1+\chi_2)} = O\left(\frac{\alpha\eta}{T\sqrt{\chi_1+\chi_2}} + \frac{\alpha^2\eta^2}{T^2|\mathcal{I}|_o(\chi_1+\chi_2)}\right) = o(\frac{N}{\eta}).$$

The second bound and Weyl's inequality yield $\psi_{\min}(\breve{\beta}'\breve{\beta}) \geq \psi_{\min}(\beta'\beta) - \left\|\breve{\beta}'\breve{\beta} - \beta'\beta\right\| \geq CN/\eta$ for some $C > 0$, implying that $\left\|(\breve{\beta}'\breve{\beta})^{-1}\right\| = O(\eta N^{-1})$. We then have

$$(\breve{\beta}'\breve{\beta})^{-1} - (\beta'\beta)^{-1} = (\beta'\beta)^{-1}(\beta'\beta - \breve{\beta}'\breve{\beta})(\breve{\beta}'\breve{\beta})^{-1} = O(\frac{\eta^2}{N^2})O\left(\frac{\alpha\eta}{T\sqrt{\chi_1+\chi_2}} + \frac{\alpha^2\eta^2}{T^2|\mathcal{I}|_o(\chi_1+\chi_2)}\right).$$

As a result, we have

$$\begin{aligned}
\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}} - \bar{\breve{\beta}}_{\mathcal{I}}'(\breve{\beta}'\breve{\beta})^{-1}\bar{\breve{\beta}}_{\mathcal{I}} &= \bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}} - \bar{\beta}_{\mathcal{I}}'(\breve{\beta}'\breve{\beta})^{-1}\bar{\beta}_{\mathcal{I}} + \bar{\beta}_{\mathcal{I}}'(\breve{\beta}'\breve{\beta})^{-1}\bar{\beta}_{\mathcal{I}} - \bar{\breve{\beta}}_{\mathcal{I}}'(\breve{\beta}'\breve{\beta})^{-1}\bar{\breve{\beta}}_{\mathcal{I}} \\
&= O\left(\frac{\alpha\eta^3}{N^2T\sqrt{\chi_1+\chi_2}} + \frac{\alpha^2\eta^4}{N^2T^2|\mathcal{I}|_o(\chi_1+\chi_2)}\right) + (\bar{\beta}_{\mathcal{I}} - \bar{\breve{\beta}}_{\mathcal{I}})'(\breve{\beta}'\breve{\beta})^{-1}\bar{\beta}_{\mathcal{I}} + \bar{\breve{\beta}}_{\mathcal{I}}'(\breve{\beta}'\breve{\beta})^{-1}(\bar{\beta}_{\mathcal{I}} - \bar{\breve{\beta}}_{\mathcal{I}}) \\
&= o(\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}}) + O\left(\frac{\alpha}{N|\mathcal{I}|_o T}\frac{\eta^2}{\sqrt{\chi_1+\chi_2}}\right) = o(\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}})
\end{aligned}$$

due to Assumption 3.1 and 3.4. Similarly, we can have that $\bar{F}_{\mathcal{T}}'(F'F)^{-1}\bar{F}_{\mathcal{T}} - \bar{\breve{F}}_{\mathcal{T}}'(\breve{F}'\breve{F})^{-1}\bar{\breve{F}}_{\mathcal{T}} = o(\bar{F}_{\mathcal{T}}'(F'F)^{-1}\bar{F}_{\mathcal{T}})$.

(ii) A little algebra reveals that

$$\|\Delta\|_F^2 = \left\|\beta(\breve{F}-F)' + (\breve{\beta}-\beta)F'\right\|_F^2 + 2\langle\beta(\breve{F}-F)' + (\breve{\beta}-\beta)F', (\breve{\beta}-\beta)(\breve{F}-F)'\rangle + \left\|(\breve{\beta}-\beta)(\breve{F}-F)'\right\|_F^2.$$

The first term leads to

$$\begin{aligned}
\left\|\beta(\breve{F}-F)' + (\breve{\beta}-\beta)F'\right\|_F^2 &= \left\|\beta(\breve{F}-F)'\right\|_F^2 + \left\|(\breve{\beta}-\beta)F'\right\|_F^2 + 2\langle\beta(\breve{F}-F)', (\breve{\beta}-\beta)F'\rangle \\
&= \frac{\alpha^2}{\chi_1+\chi_2}\frac{\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-1}\bar{\beta}_{\mathcal{I}}}{|\mathcal{T}|_o} + \frac{\alpha^2}{\chi_1+\chi_2}\frac{\bar{F}_{\mathcal{T}}'(F'F)^{-1}\bar{F}_{\mathcal{T}}}{|\mathcal{I}|_o} + 2\langle\beta(\breve{F}-F)', (\breve{\beta}-\beta)F'\rangle \\
&= \alpha^2 + 2\langle\beta(\breve{F}-F)', (\breve{\beta}-\beta)F'\rangle.
\end{aligned}$$

Then by the construction of $\breve{\beta}$ and $\breve{F}$,

$$2\langle\beta(\breve{F}-F)', (\breve{\beta}-\beta)F'\rangle = 2\text{tr}(F'(\breve{F}-F)\beta'(\breve{\beta}-\beta)) = 2\text{tr}(\begin{pmatrix}F_{\mathcal{T}}\\ \mathbf{0}\end{pmatrix}'(\breve{F}-F)\begin{pmatrix}\beta_{\mathcal{I}}\\ \mathbf{0}\end{pmatrix}'(\breve{\beta}-\beta))$$

$$\leq 2\left\|F_{\mathcal{T}}\right\|_F\left\|\beta_{\mathcal{I}}\right\|_F \frac{\alpha^2}{\chi_1+\chi_2}\sqrt{\frac{1}{|\mathcal{T}|_o}\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-2}\bar{\beta}_{\mathcal{I}}}\sqrt{\frac{1}{|\mathcal{I}|_o}\bar{F}_{\mathcal{T}}'(F'F)^{-2}\bar{F}_{\mathcal{T}}}$$
$$=\frac{\alpha^2}{\chi_1+\chi_2}O\left(\sqrt{\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-2}\bar{\beta}_{\mathcal{I}}}\sqrt{\bar{F}_{\mathcal{T}}'(F'F)^{-2}\bar{F}_{\mathcal{T}}}\right)=\frac{\alpha^2}{\chi_1+\chi_2}O\left(\frac{\eta^2}{NT}\right)=\alpha^2 o(1).$$

We also have

$$\left\|(\breve{\beta}-\beta)(\breve{F}-F)'\right\|_F^2 \leq \left\|\breve{\beta}-\beta\right\|_F^2\left\|\breve{F}-F\right\|_F^2=\frac{\alpha^2}{\chi_1+\chi_2}\frac{1}{|\mathcal{I}|_o|\mathcal{T}|_o}\bar{F}_{\mathcal{T}}'(F'F)^{-2}\bar{F}_{\mathcal{T}}\bar{\beta}_{\mathcal{I}}'(\beta'\beta)^{-2}\bar{\beta}_{\mathcal{I}}=\alpha^2 o(1).$$

Lastly, using the above bounds, we have

$$2\langle\beta(\breve{F}-F)'+(\breve{\beta}-\beta)F',(\breve{\beta}-\beta)(\breve{F}-F)'\rangle \leq 2\left\|\beta(\breve{F}-F)'+(\breve{\beta}-\beta)F'\right\|_F\left\|\breve{\beta}-\beta\right\|_F\left\|\breve{F}-F\right\|_F=\alpha^2 o(1).$$

As a result, we have $\|\Delta\|_F^2=\alpha^2(1+o(1))$. □

F.5 Proof for Section B

First of all, by Lemma F.1, Assumption F.1 are satisfied under our assumptions.

Lemma F.1. *Assumption F.1 are satisfied under Assumptions B.1 and B.2 by setting $\mu=C\eta$ for some constant $C>0$.*

Assumption F.1.

(i) $\gamma^{\frac{3}{2}}\tilde{\vartheta}c_{\mathrm{inv}}q^{\frac{11}{2}}\mu^{\frac{3}{2}}K^2\max\{\sqrt{N\log N},\sqrt{T\log T}\}\ll p_{\min}^{\frac{1}{2}}\min\{N,T\}$,

(ii) $\sigma\gamma^{\frac{5}{2}}\vartheta c_{\mathrm{inv}}^2q^3\mu K^{\frac{3}{2}}\max\{\sqrt{N\log N},\sqrt{T\log T}\}\ll p_{\min}^{\frac{1}{2}}\psi_{\min}\min\{\sqrt{N},\sqrt{T}\}$,

(iii) $\sigma^2\gamma^{\frac{3}{2}}c_{\mathrm{inv}}q^{\frac{5}{2}}\max\{N^{\frac{3}{2}},T^{\frac{3}{2}}\}\ll p_{\min}^{\frac{1}{2}}\psi_{\min}^2\min\{\sqrt{\log N},\sqrt{\log T}\}$,

(iv) $\max_{i,t}\left|M_{it}^R\right|\ll\frac{p_{\min}^{\frac{1}{2}}}{\min\{|\mathcal{I}|_o^{1/2},|\mathcal{T}|_o^{1/2}\}q^2K^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}+\frac{\psi_{\min}p_{\min}^{\frac{1}{2}}}{\min\{|\mathcal{I}|_o^{1/2},|\mathcal{T}|_o^{1/2}\}p_{\max}^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}\sqrt{NT}}$

Proof of Proposition B.1.

Consider the following decomposition which was used in the proof of 3.1:

$$\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\widehat{\beta}_i'\widehat{F}_t-\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}M_{it}$$
$$=\frac{1}{N|\mathcal{T}|_o}\sum_{j=1}^N\sum_{t\in\mathcal{T}}\bar{\beta}_{\mathcal{I}}'\left(\frac{1}{N}\sum_{j=1}^N\omega_{jt}\beta_j\beta_j'\right)^{-1}\beta_j\omega_{jt}\varepsilon_{jt}+\frac{1}{T|\mathcal{I}|_o}\sum_{i\in\mathcal{I}}\sum_{s=1}^T\bar{F}_{\mathcal{T}}'\left(\sum_{s=1}^T\omega_{is}F_sF_s'\right)^{-1}F_s\omega_{is}\varepsilon_{is}+\sum_{d=1}^7\mathcal{R}_{d,\mathcal{G}},$$
$$\mathcal{R}_{1,\mathcal{G}}:=\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\beta_i'H_2^{-1}R_t^F,\quad\mathcal{R}_{2,\mathcal{G}}:=\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}R_i^{\beta\prime}H_2F_t,$$
$$\mathcal{R}_{3,\mathcal{G}}:=\frac{1}{T|\mathcal{I}|_o}\sum_{i\in\mathcal{I}}\left(\sum_{s=1}^T\omega_{is}F_s\varepsilon_{is}\right)'\left(\frac{1}{T}\sum_{s=1}^T\omega_{is}F_sF_s'\right)^{-1}\frac{1}{N|\mathcal{T}|_o}\sum_{t\in\mathcal{T}}\left(\frac{1}{N}\sum_{j=1}^N\omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N\omega_{jt}\beta_j\varepsilon_{jt}\right),$$
$$\mathcal{R}_{4,\mathcal{G}}:=\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\left(\sum_{s=1}^T\omega_{is}F_s\varepsilon_{is}\right)'\left(\sum_{s=1}^T\omega_{is}F_sF_s'\right)^{-1}H_2^{-1}R_t^F,$$
$$\mathcal{R}_{5,\mathcal{G}}:=\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}R_i^{\beta\prime}H_2\left(\sum_{j=1}^N\omega_{jt}\beta_j\beta_j'\right)^{-1}\left(\sum_{j=1}^N\omega_{jt}\beta_j\varepsilon_{jt}\right),\quad\mathcal{R}_{6,\mathcal{G}}:=\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}R_i^{\beta\prime}R_t^F,\quad\mathcal{R}_{7,\mathcal{G}}:=-\mathcal{V}_{\mathcal{G}}^{-\frac{1}{2}}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}M_{it}^R.$$

Note that

$$\frac{1}{N|\mathcal{T}|_o}\sum_{j=1}^{N}\sum_{t\in\mathcal{T}}\bar{\beta}_{\mathcal{I}}'\left(\frac{1}{N}\sum_{j=1}^{N}\omega_{jt}\beta_j\beta_j'\right)^{-1}\beta_j\omega_{jt}\varepsilon_{jt}=\frac{1}{|\mathcal{T}|_o}\sum_{j=1}^{N}\sum_{t\in\mathcal{T}}\bar{U}_{\mathcal{I}}'\left(\sum_{j=1}^{N}\omega_{jt}U_jU_j'\right)^{-1}U_j\omega_{jt}\varepsilon_{jt}.$$

In addition, a simple calculation shows that

$$\sum_{j=1}^{N}\sum_{t\in\mathcal{T}}\mathbb{E}\left[\left(\bar{U}_{\mathcal{I}}'\left(\sum_{j=1}^{N}\omega_{jt}U_jU_j'\right)^{-1}U_j\omega_{jt}\varepsilon_{jt}\right)^2\middle|\mathcal{M},\Omega\right]\le\sigma^2\bar{U}_{\mathcal{I}}'\left(\sum_{j=1}^{N}\omega_{jt}U_jU_j'\right)^{-1}\bar{U}_{\mathcal{I}}\lesssim\sigma^2\frac{\mu K|\mathcal{T}|_o}{p_{\min}N}.$$

Hence, by the Bernstein inequality, we have with probability at least $1-\min\{N^{-3},T^{-3}\}$ that

$$\frac{1}{N|\mathcal{T}|_o}\sum_{j=1}^{N}\sum_{t\in\mathcal{T}}\bar{\beta}_{\mathcal{I}}'\left(\frac{1}{N}\sum_{j=1}^{N}\omega_{jt}\beta_j\beta_j'\right)^{-1}\beta_j\omega_{jt}\varepsilon_{jt}\lesssim\frac{\sigma\sqrt{\mu K}\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^{\frac{1}{2}}\sqrt{N|\mathcal{T}|_o}}.$$

In the same way, we have

$$\frac{1}{T|\mathcal{I}|_o}\sum_{i\in\mathcal{I}}\sum_{s=1}^{T}\bar{F}_{\mathcal{T}}'\left(\sum_{s=1}^{T}\omega_{is}F_sF_s'\right)^{-1}F_s\omega_{is}\varepsilon_{is}\lesssim\frac{\sigma\sqrt{\mu K}\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^{\frac{1}{2}}\sqrt{T|\mathcal{I}|_o}}.$$

Next, we bound the residual terms. By the simple extension of Proposition E.1, we have with probability at least $1-\min\{N^{-3},T^{-3}\}$ that

$$\begin{aligned}\|\mathcal{R}_{1,\mathcal{G}}\|&\le\max_i\left\|\beta_i'H_2^{-1}\right\|\max_t\left\|R_t^F\right\|\\&\lesssim\frac{\sigma p_{\max}^{\frac{3}{2}}\vartheta c_{\mathrm{inv}}q^{\frac{11}{2}}\mu^2K^3\max\{\sqrt{\log N},\sqrt{\log T}\}}{\min\{N,T\}}+\frac{\sigma^2}{\psi_{\min}}\frac{p_{\max}^{\frac{5}{2}}\vartheta c_{\mathrm{inv}}q^3\mu^{\frac{3}{2}}K^{\frac{5}{2}}\max\{\sqrt{N\log N},\sqrt{T\log T}\}}{\min\{\sqrt{N},\sqrt{T}\}}\\&\quad+\frac{\sigma^3}{\psi_{\min}^2}\frac{p_{\max}^{\frac{3}{2}}c_{\mathrm{inv}}q^{\frac{5}{2}}\mu^{\frac{1}{2}}K\max\{N,T\}}{p_{\min}^3}+\frac{p_{\max}^{\frac{1}{2}}}{p_{\min}}\max_{it}\left|M_{it}^R\right|.\end{aligned}$$

By the simple extension of Proposition F.1, we have with probability at least $1-\min\{N^{-3},T^{-3}\}$ that

$$\begin{aligned}\|\mathcal{R}_{2,\mathcal{G}}\|&\le\max_i\left\|R_i^\beta\right\|\max_t\|H_2F_t\|\\&\lesssim\frac{\sigma p_{\max}^{\frac{5}{2}}\vartheta c_{\mathrm{inv}}q^{\frac{15}{2}}\mu^3K^4\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^4\min\{N,T\}}+\frac{\sigma^2}{\psi_{\min}}\frac{p_{\max}^{\frac{7}{2}}\tilde{\vartheta}c_{\mathrm{inv}}^2q^7\mu^{\frac{5}{2}}K^{\frac{7}{2}}\max\{\sqrt{N}\log N,\sqrt{T}\log T\}}{p_{\min}^5\min\{\sqrt{N},\sqrt{T}\}}\\&\quad+\frac{\sigma^3}{\psi_{\min}^2}\frac{p_{\max}^{\frac{5}{2}}\vartheta c_{\mathrm{inv}}^2q^6\mu^{\frac{3}{2}}K^{\frac{5}{2}}\max\{N\sqrt{\log N},T\sqrt{\log T}\}}{p_{\min}^4}+\frac{\sigma^4}{\psi_{\min}^3}\frac{p_{\max}^2c_{\mathrm{inv}}q^{\frac{7}{2}}\mu^{\frac{1}{2}}K\max\{N,T\}\sqrt{NT}}{p_{\min}^4}\\&\quad+q^2\mu^{\frac{1}{2}}K\max_{it}\left|M_{it}^R\right|+\frac{\sigma p_{\max}^{\frac{1}{2}}q\mu^{\frac{1}{2}}K^{\frac{1}{2}}\sqrt{NT}}{p_{\min}\psi_{\min}}\max_{it}\left|M_{it}^R\right|.\end{aligned}$$

In addition, $\mathcal{R}_{3,\mathcal{G}}$ - $\mathcal{R}_{7,\mathcal{G}}$ are dominated by the above terms. Hence, we have with probability at least $1-\min\{N^{-3},T^{-3}\}$ that

$$\begin{aligned}\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\left(\widehat{M}_{it}-M_{it}\right)&\lesssim\frac{\sigma\sqrt{\mu K}\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^{\frac{1}{2}}\sqrt{N|\mathcal{T}|_o}}+\frac{\sigma\sqrt{\mu K}\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^{\frac{1}{2}}\sqrt{T|\mathcal{I}|_o}}\\&\quad+\frac{\sigma p_{\max}^{\frac{5}{2}}\vartheta c_{\mathrm{inv}}q^{\frac{15}{2}}\mu^3K^4\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^4\min\{N,T\}}+\frac{\sigma^2}{\psi_{\min}}\frac{p_{\max}^{\frac{7}{2}}\tilde{\vartheta}c_{\mathrm{inv}}^2q^7\mu^{\frac{5}{2}}K^{\frac{7}{2}}\max\{\sqrt{N}\log N,\sqrt{T}\log T\}}{p_{\min}^5\min\{\sqrt{N},\sqrt{T}\}}\\&\quad+\frac{\sigma^3}{\psi_{\min}^2}\frac{p_{\max}^{\frac{5}{2}}\vartheta c_{\mathrm{inv}}^2q^6\mu^{\frac{3}{2}}K^{\frac{5}{2}}\max\{N\sqrt{\log N},T\sqrt{\log T}\}}{p_{\min}^4}+\frac{\sigma^4}{\psi_{\min}^3}\frac{p_{\max}^2c_{\mathrm{inv}}q^{\frac{7}{2}}\mu^{\frac{1}{2}}K\max\{N,T\}\sqrt{NT}}{p_{\min}^4}.\end{aligned}$$

The terms including $\max_{it}|M_{it}^R|$ are bounded due to Assumption F.1 (iv). Lastly, by using the facts $q \lesssim K^\delta$, $\mu \lesssim \eta$, $c_{\text{inv}} \lesssim K^g \frac{\psi_{\min}}{\sqrt{NT}}$, $\psi_{NT} \lesssim \psi_{\min}$, and $q\psi_{\min} \lesssim \sqrt{NT}$ proved in the proof of Lemma E.4, we have with probability at least $1-\min\{N^{-3}, T^{-3}\}$

$$\left\|\frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}}\widehat{M}_{it} - \frac{1}{|\mathcal{G}|_o}\sum_{(i,t)\in\mathcal{G}} M_{it}\right\| \le C\left(\frac{\sigma\eta^{\frac{1}{2}}K^{\frac{1}{2}}\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^{\frac{1}{2}}\sqrt{N|\mathcal{T}|_o},} + \frac{\sigma\eta^{\frac{1}{2}}K^{\frac{1}{2}}\max\{\sqrt{\log N},\sqrt{\log T}\}}{p_{\min}^{\frac{1}{2}}\sqrt{T|\mathcal{I}|_o}} + \frac{\sigma\tilde{\vartheta}\gamma^{\frac{7}{2}}K^{(4+2g+\frac{13}{2}\delta)}\eta^3\max\{\log N,\log T\}}{p_{\min}^{\frac{3}{2}}\min\{N,T\}} + \frac{\sigma^3\gamma^2K^{(\frac{7}{2}\delta+g+1)}\eta^{\frac{1}{2}}\max\{N,T\}}{p_{\min}^2\psi_{NT}^2}\right)$$

for some constant $C>0$. □

Proof of Lemma F.1.

By using the facts $q \lesssim K^\delta$, $\mu \lesssim \eta$, $c_{\text{inv}} \lesssim K^g \frac{\psi_{\min}}{\sqrt{NT}}$, $\psi_{NT} \lesssim \psi_{\min}$, and $q\psi_{\min} \lesssim \sqrt{NT}$ proved in the proof of Lemma E.4, we can get the result. Since the proof is similar to that of Lemma E.4, we omit it. □

G Proof of Lemma E.2

In this section, we provide the remaining proof of Lemma E.2, i.e., proofs of Lemma G.1-G.5. First, we introduce several notations which are used in the proof. For any scalars a and b, $a \lesssim b$ means $|a|/|b| \le C$ for some constant $C>0$. $a \overset{c}{\asymp} b$ $c_2|b| \le |a| \le c_1|b|$ for some constants $c_1, c_2 > 0$. We write $a \ll b$ to indicate that $|a| \le c_1|b|$ for some sufficiently small constant $c_1 > 0$ and use $a \gg b$ to indicate that $c_2|a| \ge |b|$ for some sufficiently small constant $c_2>0$. For any vector a, $\|a\|_2$ denote the l_2 norm. The following five lemmas collectively prove Lemma E.2.

In Section G, we assume the following conditions:

$$\text{(i) } \sigma\tilde{\vartheta}^{\frac{1}{2}}p_{\max}^{\frac{1}{2}}q^2\mu^{\frac{1}{2}}K^{\frac{1}{2}}\max\{N\sqrt{\log N}, T\sqrt{\log T}\} \ll p_{\min}\psi_{\min}\max\{\sqrt{N},\sqrt{T}\},$$
$$\text{(ii) } \tilde{\vartheta}p_{\max}^{\frac{1}{2}}q^2\mu K\max\{\sqrt{N\log N},\sqrt{T\log T}\} \ll p_{\min}\max\{N,T\}, \quad \text{(iii) } \max_{it}|M_{it}^R| \ll \frac{1}{\max\{\sqrt{N},\sqrt{T}\}},$$

where $\tilde{\vartheta} = \max\{\vartheta, \log N + \log T\}$.

Lemma G.1. *Set $\lambda = C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ for some large constant $C_\lambda > 0$. Then, with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$,*

$$\max_t\left\|\widetilde{W}^{[t+N]}\widetilde{H}^{[t+N]} - \breve{W}^{(t+N)}\breve{H}^{(t+N)}\right\|_F \le C\frac{\sigma p_{\max}^{\frac{1}{2}}\vartheta^{\frac{1}{2}}q^{\frac{3}{2}}\mu^{\frac{1}{2}}K^{\frac{1}{2}}\max\{\sqrt{N\log N},\sqrt{T\log T}\}}{p_{\min}\psi_{\min}^{1/2}\min\{\sqrt{N},\sqrt{T}\}} \tag{G.1}$$

Proof. This is due to Lemma G.16 and (G.7). □

Lemma G.2. *Set $\lambda = C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ for some large constant $C_\lambda > 0$. Then, with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$,*

$$\max_t\left\|\widetilde{W}^{[t+N]}\widetilde{H}^{[t+N]} - W\right\| \le C\frac{\sigma p_{\max}^{\frac{1}{2}}q^{\frac{1}{2}}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}^{\frac{1}{2}}}. \tag{G.2}$$

Proof. This is due to Lemma G.16. □

Lemma G.3. *Set* $\lambda = C_\lambda \sigma \sqrt{\frac{\max\{N,T\} p_{\max}}{p_{\min}^2}}$ *for some large constant* $C_\lambda > 0$. *Then, with probability at least* $1 - O(\min\{N^{-4}, T^{-4}\})$,

$$\max_t \left\| \widetilde{W}^{[t+N]} \widetilde{Z}^{[t+N]\prime} - \widetilde{M} \right\|_F \le C \frac{\sigma p_{\max} \vartheta^{\frac{1}{2}} q^{\frac{7}{2}} K^{\frac{1}{2}} \max\{\sqrt{N \log N}, \sqrt{T \log T}\}}{p_{\min}^2 \min\{\sqrt{N}, \sqrt{T}\}}. \tag{G.3}$$

Proof. First, we consider the case of τ_l^*. By Lemma G.6, we have with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$, $\max_t \left\| \nabla f^{\text{infs}}(\widetilde{W}^{[t+N]}, \widetilde{Z}^{[t+N]}) \right\|_F \le c \frac{\sqrt{c_{\text{inj}} p_{\min}}}{q} \lambda \psi_{\min}^{1/2}$ where the constant $c > 0$ is defined in Assumption (G.14) of Lemma G.12. Hence, with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$, for all t, $\left(\widetilde{W}^{[t+N]} \widetilde{H}^{[t+N]}, \widetilde{Z}^{[t+N]} \widetilde{H}^{[t+N]} \right)$ satisfies Assumption (G.14) of Lemma G.12. In addition, by Lemma G.16, (G.29) holds for all τ_l^* where $1 \le l \le N + T$. Hence, $\max_t \left\| \widetilde{W}^{[t+N]} \widetilde{Z}^{[t+N]\prime} - M^\star \right\| \le C q \lambda$ for some constant $C > 0$. Hence, with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$, for all t, $\left(\widetilde{W}^{[t+N]} \widetilde{H}^{[t+N]}, \widetilde{Z}^{[t+N]} \widetilde{H}^{[t+N]} \right)$ satisfies assumptions of Lemma G.12. Similarly, we can show that for all t, $\left(\widetilde{W}^{[t+N]} \widetilde{H}^{[t+N]}, \widetilde{Z}^{[t+N]} \widetilde{H}^{[t+N]} \right)$ satisfies assumptions of Lemma G.12. Therefore, by Lemma G.12, we have $\max_t \left\| \widetilde{W}^{[t+N]} \widetilde{Z}^{[t+N]\prime} - \widetilde{M} \right\|_F \le C \frac{\vartheta^{\frac{1}{2}} \sigma p_{max}}{p_{\min}^2} \frac{q^{\frac{5}{2}} \mu^{\frac{1}{2}} K \max\{\sqrt{N \log N}, \sqrt{T \log T}\}}{c_{\text{inj}} \min\{\sqrt{N}, \sqrt{T}\}}$ for some constant $C > 0$. Then, by setting $c_{\text{inj}} = 1/(32q)$, we have the desired result. Second, we consider the case of τ^*. The proof for this case is analogous to the case of τ_l^*. The only difference is that we use Lemma G.23 in lieu of Lemma G.6. In this way, we can obtain the same bound as τ_l^* case. □

Lemma G.4. *Set* $\lambda = C_\lambda \sigma \sqrt{\frac{\max\{N,T\} p_{\max}}{p_{\min}^2}}$ *for some large constant* $C_\lambda > 0$. *Then, with probability at least* $1 - O(\min\{N^{-4}, T^{-4}\})$,

$$\left\| \widetilde{M} - M^\star \right\| \le C \frac{\sigma p_{\max}^{\frac{1}{2}} q \max\{\sqrt{N}, \sqrt{T}\}}{p_{\min}}. \tag{G.4}$$

Proof. Set any $1 \le t^* \le T$. Note that $\left\| \widetilde{M} - M^\star \right\| \le \left\| \widetilde{M} - \widetilde{W}^{[t^*+N]} \widetilde{Z}^{[t^*+N]\prime} \right\|_F + \left\| \widetilde{W}^{[t^*+N]} \widetilde{Z}^{[t^*+N]\prime} - M^\star \right\|$. Then, Lemma G.16 and Lemma G.3 yield the desired bound.

□

Lemma G.5. *With probability at least* $1 - O(\min\{N^{-5}, T^{-5}\})$,

$$\max_t \left\| \breve{W}^{(t+N)} \breve{H}^{(t+N)} - W \right\|_{2,\infty} \le C \frac{\sigma p_{\max}^{\frac{1}{2}} \vartheta^{\frac{1}{2}} q^{\frac{3}{2}} \mu^{\frac{1}{2}} K^{\frac{1}{2}} \max\{\sqrt{N \log N}, \sqrt{T logT}\}}{p_{\min} \psi_{\min}^{1/2} \min\{\sqrt{N}, \sqrt{T}\}}. \tag{G.5}$$

Proof. This is due to Lemma G.16. □

G.1 Lemmas regarding the new stopping point

We now introduce Lemma G.6 which provides a uniform bound for gradients of (D.1). Together with Lemma G.12 and Lemma G.16, it guarantees that our non-convex estimators at the new stopping points, $\widetilde{W}^{[t+N]} \widetilde{Z}^{[t+N]\prime}$, closely approximate the convex estimator, $\widetilde{M}$, uniformly.

Lemma G.6. *Recall that* $\tau_l^* = \arg\min_{0 \le \tau \le \tau_0} \left\| \nabla f^{\text{infs},(l)}(W^{\tau,(l)}, Z^{\tau,(l)}) \right\|_F$ *for each* $1 \le l \le N + T$. *With probability at least* $1 - O(\min\{N^{-5}, T^{-5}\})$, *we have*

$$\max_{1 \le l \le N+T} \left\| \nabla f^{\text{infs}}(W^{\tau_l^*}, Z^{\tau_l^*}) \right\|_F \le C \left(\frac{\sigma \max\{\sqrt{N \log N}, \sqrt{T \log T}\}}{p_{\min}} + \lambda \right) \frac{\vartheta^{\frac{1}{2}} q \mu^{\frac{1}{2}} K^{\frac{3}{2}} \psi_{\min}^{\frac{1}{2}}}{\min\{\sqrt{N}, \sqrt{T}\}}.$$

Proof. By Lemma G.8, with probability at least $1-O(\min\{N^{-5},T^{-5}\})$, we have $\max_{1\le l\le N+T}\left\|\nabla f^{\text{infs},(l)}(W^{\tau_l^*,(l)},Z^{\tau_l^*,(l)})\right\|_F \le C_{gr}\frac{1}{\max\{N^5,T^5\}}\lambda\sqrt{\psi_{\min}}$. In addition, by Lemma G.7, with probability at least $1-O(\min\{N^{-5},T^{-5}\})$, (G.8) holds for all $0\le\tau\le\tau_0$. So, with probability at least $1-O(\min\{N^{-5},T^{-5}\})$, (G.8) holds for all τ_l^* where $1\le l\le N+T$. Therefore, we have with probability at least $1-O(\min\{N^{-5},T^{-5}\})$,

$$
\begin{aligned}
&\max_{1\le l\le N+T}\left\|\nabla f^{\text{infs}}(W^{\tau_l^*},Z^{\tau_l^*})\right\|_F\\
&\le \max_{1\le l\le N+T}\left|\left\|\nabla f^{\text{infs}}(W^{\tau_l^*},Z^{\tau_l^*})\right\|_F-\left\|\nabla f^{\text{infs},(l)}(W^{\tau_l^*,(l)},Z^{\tau_l^*,(l)})\right\|_F\right|+\max_{1\le l\le N+T}\left\|\nabla f^{\text{infs},(l)}(W^{\tau_l^*,(l)},Z^{\tau_l^*,(l)})\right\|_F\\
&\le C\left(\frac{\sigma\max\{\sqrt{N\log N},\sqrt{T\log T}\}}{p_{\min}}+\lambda\right)\frac{\sqrt{\vartheta}q\mu^{\frac12}K^{\frac32}\psi_{\min}^{\frac12}}{\min\{\sqrt{N},\sqrt{T}\}}
\end{aligned}
$$

for some constant $C>0$. □

Hereinafter, for notational simplicity, we denote $\mathcal{F}^\tau := \begin{bmatrix}W^\tau\\ Z^\tau\end{bmatrix}\in\mathbb{R}^{(N+T)\times K}$ and $\mathcal{F} := \begin{bmatrix}W\\ Z\end{bmatrix}\in\mathbb{R}^{(N+T)\times K}$. Note that we have the following properties of $\mathcal{F}$.

$$
\psi_1(\mathcal{F})=\|\mathcal{F}\|=\sqrt{2\psi_{\max}},\qquad \psi_K(\mathcal{F})=\sqrt{2\psi_{\min}}, \tag{G.6}
$$

$$
\|\mathcal{F}\|_{2,\infty}=\max\{\|W\|_{2,\infty},\|Z\|_{2,\infty}\}\le\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}. \tag{G.7}
$$

Lemma G.7. *With probability at least* $1-O(\min\{N^{-5},T^{-5}\})$*, we have*

$$
\begin{aligned}
&\max_{0\le\tau\le\tau_0}\max_{1\le l\le N+T}\left|\left\|\nabla f^{\text{infs}}(W^\tau,Z^\tau)\right\|_F-\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)},Z^{\tau,(l)})\right\|_F\right|\\
&\le C_{no}K\sqrt{\vartheta}\psi_{\max}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N,T\log T\}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}
\end{aligned} \tag{G.8}
$$

for some constant $C_{no}>0$.

Proof. Pick arbitrary $\tau\le\tau_0$, and suppose that (G.16)-(G.21) hold at the τth iteration. Without loss of generality, we consider the case when $l\le N$. The other case can be shown similarly. By the definition of $f^{\text{infs}}(\cdot,\cdot)$ and $f^{\text{infs},(l)}(\cdot,\cdot)$, we have that $\nabla f^{\text{infs}}(\mathcal{A}O)=\nabla f^{\text{infs}}(\mathcal{A})O$ and $\nabla f^{\text{infs},(l)}(\mathcal{A}O)=\nabla f^{\text{infs},(l)}(\mathcal{A})O$ for any $\mathcal{A}\in\mathbb{R}^{(N+T)\times K}$, $O\in\mathcal{O}^{K\times K}$. Then, using the unitary invariance of $\|\cdot\|_F$, we can write

$$
\begin{aligned}
&\left|\left\|\nabla f^{\text{infs}}(W^\tau,Z^\tau)\right\|_F-\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)},Z^{\tau,(l)})\right\|_F\right|=\left|\left\|\nabla f^{\text{infs}}(W^\tau H^\tau,Z^\tau H^\tau)\right\|_F-\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)}Q^{\tau,(l)},Z^{\tau,(l)}Q^{\tau,(l)})\right\|_F\right|\\
&\le \underbrace{\left\|\nabla f^{\text{infs}}(W^\tau H^\tau,Z^\tau H^\tau)-\nabla f^{\text{infs}}(W^{\tau,(l)}Q^{\tau,(l)},Z^{\tau,(l)}Q^{\tau,(l)})\right\|_F}_{:=\vartheta_1}\\
&\quad+\underbrace{\left\|\nabla f^{\text{infs}}(W^{\tau,(l)}Q^{\tau,(l)},Z^{\tau,(l)}Q^{\tau,(l)})-\nabla f^{\text{infs},(l)}(W^{\tau,(l)}Q^{\tau,(l)},Z^{\tau,(l)}Q^{\tau,(l)})\right\|_F}_{:=\vartheta_2}.
\end{aligned}
$$

First, we control ϑ_1. Note that $\vartheta_1=\left\|\begin{bmatrix}\vartheta_3\\ \vartheta_4\end{bmatrix}\right\|_F$ where

$$
\begin{aligned}
\vartheta_3&=\Pi^{-1}\mathcal{P}_\Omega(W^\tau Z^{\tau\prime}-Y)Z^\tau H^\tau+\lambda W^\tau H^\tau-\Pi^{-1}\mathcal{P}_\Omega(W^{\tau,(l)}Z^{\tau,(l)\prime}-Y)Z^{\tau,(l)}Q^{\tau,(l)}-\lambda W^{\tau,(l)}Q^{\tau,(l)}\\
\vartheta_4&=\left(\Pi^{-1}\mathcal{P}_\Omega(W^\tau Z^{\tau\prime}-Y)\right)'W^\tau H^\tau+\lambda Z^\tau H^\tau-\left(\Pi^{-1}\mathcal{P}_\Omega(W^{\tau,(l)}Z^{\tau,(l)\prime}-Y)\right)'W^{\tau,(l)}Q^{\tau,(l)}-\lambda Z^{\tau,(l)}Q^{\tau,(l)}
\end{aligned}
$$

Rearrange ϑ_3 to obtain

$$\vartheta_3 = \underbrace{\Pi^{-1}\mathcal{P}_\Omega(W^\tau Z^{\tau\prime})Z^\tau H^\tau - \Pi^{-1}\mathcal{P}_\Omega(W^{\tau,(l)}Z^{\tau,(l)\prime})Z^{\tau,(l)}Q^{\tau,(l)}}_{:=\beta_1}$$
$$- \underbrace{\Pi^{-1}\mathcal{P}_\Omega(M^\star)(Z^\tau H^\tau - Z^{\tau,(l)}Q^{\tau,(l)})}_{:=\beta_2} - \underbrace{\Pi^{-1}\mathcal{P}_\Omega(M^R+\mathcal{E})(Z^\tau H^\tau - Z^{\tau,(l)}Q^{\tau,(l)})}_{:=\beta_3} + \underbrace{\lambda W^\tau H^\tau - \lambda W^{\tau,(l)}Q^{\tau,(l)}}_{:=\beta_4}$$

For $\|\beta_1\|_F$, we have

$$\begin{aligned}
\|\beta_1\|_F &= \left\|\Pi^{-1}\mathcal{P}_\Omega(W^\tau Z^{\tau\prime})(Z^\tau H^\tau - Z^{\tau,(l)}Q^{\tau,(l)}) + \Pi^{-1}\mathcal{P}_\Omega(W^\tau Z^{\tau\prime} - W^{\tau,(l)}Z^{\tau,(l)\prime})Z^{\tau,(l)}Q^{\tau,(l)}\right\|_F \\
&\leq \left\|\Pi^{-1}\mathcal{P}_\Omega(W^\tau Z^{\tau\prime})(Z^\tau H^\tau - Z^{\tau,(l)}Q^{\tau,(l)})\right\|_F \\
&\quad + \left\|\Pi^{-1}\mathcal{P}_\Omega((W^\tau H^\tau - W^{\tau,(l)}Q^{\tau,(l)})(Z^\tau H^\tau)' - W^{\tau,(l)}Q^{\tau,(l)}(Z^\tau H^\tau - Z^{\tau,(l)}Q^{\tau,(l)})')Z^{\tau,(l)}Q^{\tau,(l)}\right\|_F \\
&\lesssim p_{\min}^{-1}\|W\|_F^2\left\|\mathcal{F}^\tau H^\tau - \mathcal{F}^{\tau,(l)}Q^{\tau,(l)}\right\|_F \lesssim K\psi_{\max}\sqrt{\vartheta}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}
\end{aligned}$$

where the third relation uses the unitary invariance of the Frobenius norm, the fourth relation follows from Lemma G.22 (ii) and (iv). The last relation comes from (G.7) and (G.18).

For $\|\beta_2\|_F$, we have

$$\begin{aligned}
\|\beta_2\|_F &\leq \left\|\Pi^{-1}\right\|\|\mathcal{P}_\Omega(M^\star)\|_F\left\|\mathcal{F}^\tau H^\tau - \mathcal{F}^{\tau,(l)}Q^{\tau,(l)}\right\|_F \\
&\lesssim \sqrt{K}\psi_{\max}\sqrt{\vartheta}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}
\end{aligned}$$

where the second line comes from (G.7) and (G.18). For $\|\beta_3\|_F$, Lemma G.13, Lemma G.14, (G.7) and (G.18) yield that, with probability at least $1 - O(\min\{N^{-100}, T^{-100}\})$,

$$\begin{aligned}
\|\beta_3\|_F &\leq \left\|\Pi^{-1}\right\|\left(\left\|\mathcal{P}_\Omega(M^R)\right\| + \|\mathcal{P}_\Omega(\mathcal{E})\|\right)\left\|\mathcal{F}^\tau H^\tau - \mathcal{F}^{\tau,(l)}Q^{\tau,(l)}\right\|_F \\
&\lesssim \sigma\sqrt{\vartheta}\sqrt{\max\{N,T\}p_{\max}}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}.
\end{aligned}$$

For $\|\beta_4\|_F$, (G.7) and (G.18) yield $\|\beta_4\|_F \lesssim \lambda\sqrt{\vartheta}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}$. Combining the above bounds, we reach

$$\|\vartheta_3\|_F \leq \|\beta_1\|_F + \|\beta_2\|_F + \|\beta_3\|_F + \|\beta_4\|_F \lesssim K\sqrt{\vartheta}\psi_{\max}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}.$$

ϑ_4 can be bounded in an analogous way, and thus we have $\vartheta_1 \lesssim K\sqrt{\vartheta}\psi_{\max}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}$.

We can bound ϑ_2 in an analogous fashion as the the proof for bounding A_3 in Lemma 12 of Chen et al. (2020). Additionally, use Lemma G.22 and (G.7) to obtain $\vartheta_2 \lesssim \sqrt{\vartheta}\sigma\sqrt{\frac{p_{\max}\max\{N\log N, T\log T\}}{p_{\min}^2}}\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}$ with probability at least $1 - O(\min\{N^{-100}, T^{-100}\})$. Therefore, we conclude that, with probability at least $1 - O(\min\{N^{-100}, T^{-100}\})$,

$$\left|\left\|\nabla f^{\text{infs}}(W^\tau, Z^\tau)\right\|_F - \left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)}, Z^{\tau,(l)})\right\|_F\right| \leq C_{no}K\sqrt{\vartheta}\psi_{\max}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}.$$

for some constant $C_{no} > 0$. This result and Lemma G.16 together complete the proof. □

Lemma G.8. *Suppose that $\eta \overset{c}{\precsim} 1/\max\{N,T\}q^3\psi_{\max}$. Then we have, with probability at least $1-O(\min\{N^{-5},T^{-5}\})$, there is a constant $C_{gr}>0$ such that*

$$\max_{1\le l\le N+T}\min_{0\le\tau<\tau_0}\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)},Z^{\tau,(l)})\right\|_F \le C_{gr}\frac{1}{\max\{N^5,T^5\}}\lambda\sqrt{\psi_{\min}}. \tag{G.9}$$

Proof. First, note that (G.16)-(G.21) hold for all $0\le\tau\le\tau_0$, and (G.22) hold for all $1\le\tau\le\tau_0$ and for all $1\le l\le N+T$, with probability at least $1-O(\min\{N^{-5},T^{-5}\})$, by Lemma G.16 and G.21. Without loss of generality, we consider the case when $l\le N$. The other case can be shown similarly. By (G.22), we have that $f^{\text{infs},(l)}(W^{\tau_0,(l)},Z^{\tau_0,(l)})\le f^{\text{infs},(l)}(W^{0,(l)},Z^{0,(l)})-\frac{\eta}{2}\sum_{\tau=0}^{\tau_0-1}\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)},Z^{\tau,(l)})\right\|_F^2$. Then, we have

$$\begin{aligned}\min_{0\le\tau<\tau_0}\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)},Z^{\tau,(l)})\right\|_F &\le \left\{\frac{1}{\tau_0}\sum_{\tau=0}^{\tau_0-1}\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)},Z^{\tau,(l)})\right\|_F^2\right\}^{1/2}\\ &\le \left\{\frac{2}{\eta\tau_0}\left(f^{\text{infs},(l)}(W,Z)-f^{\text{infs},(l)}(W^{\tau_0,(l)},Z^{\tau_0,(l)})\right)\right\}^{1/2}.\end{aligned} \tag{G.10}$$

Now we control $f^{\text{infs},(l)}(W,Z)-f^{\text{infs},(l)}(W^{\tau_0,(l)},Z^{\tau_0,(l)})$. Recalling that $f^{\text{infs},(l)}(A,B)=f^{\text{infs},(l)}(AO,BO)$ for any $O\in\mathcal{O}^{K\times K}$, we have

$$\begin{aligned}f^{\text{infs},(l)}(\mathcal{F}^{\tau_0,(l)})=f^{\text{infs},(l)}(\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)})=&f^{\text{infs},(l)}(\mathcal{F})+\langle\nabla f^{\text{infs},(l)}(\mathcal{F}),\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F}\rangle\\ &+\frac{1}{2}\text{vec}(\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F})'\nabla^2 f^{\text{infs},(l)}(\bar{\mathcal{F}}^{(l)})\text{vec}(\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F})\end{aligned}$$

where $\bar{\mathcal{F}}^{(l)}$ is a point between $\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}$ and $\mathcal{F}$. Before proceeding, notice that

$$\begin{aligned}\left\|\bar{\mathcal{F}}^{(l)}-\mathcal{F}\right\|_{2,\infty}&\le\left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F}\right\|_{2,\infty}\left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F}^{\tau_0}H^{\tau_0}\right\|_F+\left\|\mathcal{F}^{\tau_0}H^{\tau_0}-\mathcal{F}\right\|_{2,\infty}\\ &\lesssim\sqrt{\vartheta}q\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N,T\log T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\sqrt{\frac{\mu K\psi_{\max}}{\min\{N,T\}}}\le\frac{1}{1000q\sqrt{N+T}}\sqrt{\psi_{\max}},\end{aligned} \tag{G.11}$$

where the second line is due to the triangle inequality, the other lines use (G.18) , (G.20) and (G.7). This allows us to invoke Lemma G.9 to conclude that $\left\|\nabla^2 f^{\text{infs},(l)}(\bar{\mathcal{F}}^{(l)})\right\|\lesssim\psi_{\max}$ with probability at least $1-O(\min\{N^{-99},T^{-99}\})$. Then the triangle inequality yields

$$\begin{aligned}&f^{\text{infs},(l)}(\mathcal{F})-f^{\text{infs},(l)}(\mathcal{F}^{\tau_0,(l)})\\ &\le\left\|\nabla f^{\text{infs},(l)}(\mathcal{F})\right\|_F\left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F}\right\|_F-\frac{1}{2}\text{vec}(\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F})'\nabla^2 f^{\text{infs},(l)}(\bar{\mathcal{F}}^{(l)})\text{vec}(\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F})\\ &\le\left\|\nabla f^{\text{infs},(l)}(\mathcal{F})\right\|_F\left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F}\right\|_F+\frac{C}{2}\psi_{\max}\left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)}-\mathcal{F}\right\|_F^2,\end{aligned}$$

with probability at least $1-O(\min\{N^{-99},T^{-99}\})$, for some constant $C>0$. To bound $\left\|\nabla f^{\text{infs},(l)}(\mathcal{F})\right\|_F$, note that

$$\begin{aligned}\left\|\nabla f^{\text{infs},(l)}(\mathcal{F})\right\|_F&\le\left\|\nabla_W f^{\text{infs},(l)}(\mathcal{F})\right\|_F+\left\|\nabla_Z f^{\text{infs},(l)}(\mathcal{F})\right\|_F\\ &\le\left\|\Pi^{-1}\mathcal{P}_{\Omega_{-l,\cdot}}(M^R+\mathcal{E})Z\right\|_F+\lambda\left\|W\right\|_F+\left\|(\Pi^{-1}\mathcal{P}_{\Omega_{-l,\cdot}}(M^R+\mathcal{E}))'W\right\|_F+\lambda\left\|Z\right\|_F\\ &\le(\left\|\Pi^{-1}\mathcal{P}_{\Omega_{-l,\cdot}}(M^R+\mathcal{E})\right\|+\lambda)(\left\|W\right\|_F+\left\|Z\right\|_F)\lesssim\left(\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}+\lambda\right)\sqrt{K\psi_{\max}}\overset{c}{\precsim}\lambda\sqrt{K\psi_{\max}},\end{aligned} \tag{G.12}$$

where the penultimate relation holds by Lemmas G.13 and Lemma G.14 with probability at least $1-O(\min\{N^{-100},T^{-100}\})$, and the last relation is due to the assumption that $\lambda = C_\lambda \sigma \sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$. To bound $\left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)} - \mathcal{F}\right\|_F$, we use (G.7), (G.16), (G.18), and $\min\{N,T\} \gg \vartheta q \mu \max\{\log N, \log T\}$ to obtain

$$\left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)} - \mathcal{F}\right\|_F \leq \left\|\mathcal{F}^{\tau_0,(l)}Q^{\tau_0,(l)} - \mathcal{F}^{\tau_0}H^{\tau_0}\right\|_F + \left\|\mathcal{F}^{\tau_0}H^{\tau_0} - \mathcal{F}\right\|_F \lesssim \left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|W\|_F.$$

Using these bounds, with probability at least $1 - O(\min\{N^{-99}, T^{-99}\})$, we can write

$$\begin{aligned} f^{\text{infs},(l)}(\mathcal{F}) - f^{\text{infs},(l)}(\mathcal{F}^{\tau_0,(l)}) &\lesssim \lambda\sqrt{K\psi_{\max}}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|W\|_F \\ &\quad + \psi_{\max}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)^2 \|W\|_F^2 \lesssim Kq^2\lambda^2, \end{aligned}$$

where the first relation uses (G.16) with $\tau = \tau_0$, and the last relation comes from $\lambda = C_\lambda \sigma \sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$. Putting this back into (G.10), we reach, with probability at least $1 - O(\min\{N^{-99}, T^{-99}\})$, for some constant $C_{gr} > 0$,

$$\min_{0 \leq \tau < \tau_0} \left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)}, Z^{\tau,(l)})\right\|_F \leq C_{gr}\sqrt{\frac{1}{\eta\tau_0}Kq^2\lambda^2} \leq C_{gr}\frac{1}{\max\{N^5, T^5\}}\lambda\sqrt{\psi_{\min}},$$

as long as $\eta \overset{c}{\asymp} \frac{1}{\max\{N,T\}q^3\psi_{\max}}$, $\tau_0 = \max\{N^{18}, T^{18}\}$ and that $\max\{N,T\} \geq q$. The last is a consequence of our assumption $\max\{N,T\} \geq \min\{N,T\} \gg q\max\{\log N, \log T\}$. Therefore, with probability at least $1 - O(\min\{N^{-98}, T^{-98}\})$, we have

$$\max_{1 \leq l \leq N+T} \min_{0 \leq \tau < \tau_0} \left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)}, Z^{\tau,(l)})\right\|_F \leq C_{gr}\frac{1}{\max\{N^5, T^5\}}\lambda\sqrt{\psi_{\min}},$$

which completes the proof. □

Lemma G.9. *With probability at least $1 - O(\min\{N^{-99}, T^{-99}\})$, there exists a constant $C > 0$ such that*

$$\max_{1 \leq l \leq N+T} \left\|\nabla^2 f^{\text{infs},(l)}(A, B)\right\| \leq C\psi_{\max}$$

hold simultaneously for all $A \in \mathbb{R}^{N \times K}$, $B \in \mathbb{R}^{T \times K}$ satisfying $\left\|\begin{bmatrix} A - W \\ B - Z \end{bmatrix}\right\|_{2,\infty} \leq \frac{1}{500q\sqrt{N+T}}\|W\|$.

Proof. Without loss of generality, we consider the case when $l \leq N$. The other case can be shown similarly. Note first that, for any $\Delta \in \mathbb{R}^{(N+T)\times K}$,

$$\begin{aligned} \text{vec}(\Delta)'\nabla^2 f^{\text{infs},(l)}(A,B)\text{vec}(\Delta) &= \left\|\Pi^{-1/2}\mathcal{P}_\Omega(A\Delta_Z' + \Delta_W B')\right\|_F^2 + 2\langle \Pi^{-1}\mathcal{P}_\Omega(AB' - M^\star - M^R - \mathcal{E}), \Delta_W\Delta_Z'\rangle + \lambda\|\Delta\|_F^2 \\ &\quad - \left\|\Pi^{-1/2}\mathcal{P}_{\Omega_{l,\cdot}}(A\Delta_Z' + \Delta_W B')\right\|_F^2 - 2\langle \Pi^{-1}\mathcal{P}_{\Omega_{l,\cdot}}(AB' - M^\star - M^R - \mathcal{E}), \Delta_W\Delta_Z'\rangle \\ &\quad + \left\|\mathcal{P}_{l,\cdot}(A\Delta_Z' + \Delta_W B')\right\|_F^2 + 2\langle \mathcal{P}_{l,\cdot}(AB' - M^\star), \Delta_W\Delta_Z'\rangle \end{aligned}$$

where $\left\|\Pi^{-1/2}\mathcal{P}_\Omega(A\Delta_Z' + \Delta_W B')\right\|_F^2 + 2\langle \Pi^{-1}\mathcal{P}_\Omega(AB' - M^\star - M^R - \mathcal{E}), \Delta_W\Delta_Z'\rangle + \lambda\|\Delta\|_F^2 \lesssim \psi_{\max}\|\Delta\|_F^2$ with probability at least $1 - O(\min\{N^{-100}, T^{-100}\})$ by a simple extension of the proof of Lemma 17 of Chen et al. (2020). We would bound the remaining terms in turn. First, note that

$$\left\|\mathcal{P}_{l,\cdot}(A\Delta_Z' + \Delta_W B')\right\|_F^2 \leq \left\|A\Delta_Z' + \Delta_W B'\right\|_F^2 \leq 2\left(\|W\Delta_Z'\|_F^2 + \|\Delta_W Z'\|_F^2\right) \leq 2\left(\|W\|^2\|\Delta_Z\|_F^2 + \|Z\|^2\|\Delta_W\|_F^2\right) = 2\psi_{\max}\|\Delta\|_F^2$$

Second, $\left\|\Pi^{-1/2}\mathcal{P}_{\Omega_{l,\cdot}}(A\Delta_Z' + \Delta_W B')\right\|_F^2 \leq p_{\min}^{-1}\left\|\mathcal{P}_{\Omega_{l,\cdot}}(A\Delta_Z' + \Delta_W B')\right\|_F^2 \leq p_{\min}^{-1}\left\|A\Delta_Z' + \Delta_W B'\right\|_F^2 \leq 2p_{\min}^{-1}\psi_{\max}\|\Delta\|_F^2$

Third,

$$
\begin{aligned}
&|2\langle \mathcal{P}_{l,\cdot}(AB'-M^{\star}),\Delta_W\Delta_Z'\rangle| \leq 2\left\|\mathcal{P}_{l,\cdot}(AB'-M^{\star})\right\|_F\left\|\Delta_W\Delta_Z'\right\|_F \leq 2\left\|AB'-M^{\star}\right\|_F\left\|\Delta\right\|_F^2 \\
&\leq 2(\|(A-W)B'\|_F+\|W(B-Z)'\|_F)\left\|\Delta\right\|_F^2 \leq 4(\|A-W\|_F+\|B-Z\|_F)\left\|W\right\|\left\|\Delta\right\|_F^2 \leq 8\psi_{\max}\left\|\Delta\right\|_F^2
\end{aligned}
$$

where we use

$$
\begin{aligned}
&\left\|\begin{bmatrix}A-W\\B-Z\end{bmatrix}\right\| \leq \left\|\begin{bmatrix}A-W\\B-Z\end{bmatrix}\right\|_F \leq \sqrt{N+T}\left\|\begin{bmatrix}A-W\\B-Z\end{bmatrix}\right\|_{2,\infty} \leq \frac{1}{500q}\left\|W\right\|, \\
&\left\|B\right\| \leq \left\|B-W\right\|+\left\|W\right\| \leq \frac{1+500q}{500q}\left\|W\right\| \leq 2\left\|W\right\|
\end{aligned}
$$

in the last two inequalities. Also, $|2\langle \Pi^{-1}\mathcal{P}_{\Omega_{l,\cdot}}(AB'-M^{\star}),\Delta_W\Delta_Z'\rangle|$ can be bounded in the same way. Lastly, Lemma G.13 and G.14 give us that, with probability $1-O(\min\{N^{-100},T^{-100}\})$,

$$
|2\langle \Pi^{-1}\mathcal{P}_{\Omega_{l,\cdot}}(M^R+\mathcal{E}),\Delta_W\Delta_Z'\rangle| \leq 2\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}\left\|\Delta\right\|_F^2 \leq \frac{1}{10}\psi_{\max}\left\|\Delta\right\|_F^2
$$

where the last inequality uses the assumption $\frac{\sigma}{\psi_{\min}}\sqrt{\frac{p_{\max}\max\{N,T\}}{p_{\min}^2}} \ll 1$. Combining these bounds yields that, with probability at least $1-O(\min\{N^{-100},T^{-100}\})$, for some constant $C>0$ which does not depend on l, $\left\|\nabla^2 f^{\text{infs},(l)}(A,B)\right\| \leq C\psi_{\max}$. Therefore, with probability at least $1-O(\min\{N^{-99},T^{-99}\})$, we obtain $\max_{1\leq l\leq N+T}\left\|\nabla^2 f^{\text{infs},(l)}(A,B)\right\| \leq C\psi_{\max}$. □

G.2 Lemmas regarding the original stopping point

Lemma G.10. *Define* $\xi_t^{\tau} := \sqrt{N}\left(W^{\tau,(t+N)}H^{\tau,(t+N)}-W\right)D_{M^{\star}}^{-\frac{1}{2}}$*. Then, we have with probability at least* $1-O(\min\{N^{-5},T^{-5}\})$ *that*

$$
\max_t\max_\tau \|\xi_t^{\tau\prime}\Omega_t\mathcal{E}_t\| \leq C\left(\frac{\sigma^2 p_{\max}^{\frac{1}{2}}\vartheta^{\frac{1}{2}}q^{\frac{1}{2}}K^{\frac{1}{2}}\sqrt{N\log T}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}}\right).
$$

Proof. First, note that, with probability at least $1-O(\min\{N^{-5},T^{-5}\})$,

$$
\max_t\max_\tau \|\xi_t^{\tau}\|_{2,\infty} \leq C\left(\frac{\sigma p_{\max}^{\frac{1}{2}}\vartheta^{\frac{1}{2}}q^{\frac{3}{2}}\mu^{\frac{1}{2}}K^{\frac{1}{2}}\sqrt{N}\max\{\sqrt{N\log N},\sqrt{T\log T}\}}{p_{\min}\psi_{\min}\min\{\sqrt{N},\sqrt{T}\}}\right)
$$

for some absolute constant $C>0$ by Lemma G.16. Second, by Lemma G.16 again, we have with probability at least $1-O(\min\{N^{-5},T^{-5}\})$,

$$
\begin{aligned}
\max_t\max_\tau \|\xi_t^{\tau}\|_F &\leq \sqrt{N}\left(\max_t\max_\tau \|W^{\tau}H^{\tau}-W^{\tau,(t+N)}H^{\tau,(t+N)}\|_F+\|W-W^{\tau}H^{\tau}\|_F\right)\|D_{M^{\star}}^{-\frac{1}{2}}\| \\
&\leq C\left(\frac{\sigma p_{\max}^{\frac{1}{2}}q^{\frac{1}{2}}K^{\frac{1}{2}}\sqrt{N}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}}\right)
\end{aligned}
$$

for some absolute constant $C>0$ by Lemma G.16. For each τ and t, because ξ_t^{τ} only depends on $M^{\star}$ and Y excluding the t-th column of Y, conditioning on $\{\mathcal{M},\Omega\}$, $\{\varepsilon_{jt}\}_{j\leq N}$ are independent of ξ_t^{τ}. Hence, $\mathbb{E}\left[\varepsilon_{jt}|\mathcal{M},\Omega,\xi_t^{\tau}\right]=\mathbb{E}\left[\varepsilon_{jt}|\mathcal{M},\Omega\right]=0$ and, conditioning on $\{\mathcal{M},\Omega,\xi_t^{\tau}\}$, $\{\varepsilon_{jt}\}_{j\leq N}$ are independent across j. Then, by matrix Bernstein inequality, we have for each τ and t that $\|\xi_t^{\tau\prime}\Omega_t\mathcal{E}_t\|=\|\sum_{j=1}^N\omega_{jt}\varepsilon_{jt}\xi_{t,j}^{\tau\prime}\| \leq C\left(\sigma\log T\log N\max_t\max_\tau\|\xi_t^{\tau}\|_{2,\infty}+\sigma\sqrt{\log T}\max_t\max_\tau\|\xi_t^{\tau}\|_F\right)$ for some absolute contant $C>0$ with probability exceeding $1-O(T^{-100})$. So, we have with probability at least $1-O(\min\{N^{-5},T^{-5}\})$ that $\max_t\max_\tau\|\xi_t^{\tau\prime}\Omega_t\mathcal{E}_t\| \leq C\left(\frac{\sigma^2 p_{\max}^{\frac{1}{2}}\vartheta^{\frac{1}{2}}q^{\frac{1}{2}}K^{\frac{1}{2}}\sqrt{N\log T}\max\{\sqrt{N},\sqrt{T}\}}{p_{\min}\psi_{\min}}\right)$ for some absolute constant $C>0$. □

Lemma G.11. *With probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$,*

$$\max_t \max_\tau \| \frac{1}{\sqrt{N}} H_1' \beta' (\Omega_t - \Pi) \xi_t^\tau \| \leq C \left(\frac{\sigma p_{\max} \vartheta q^{\frac{1}{2}} \mu^{\frac{1}{2}} K \sqrt{\log T} \max\{\sqrt{N}, \sqrt{T}\}}{p_{\min} \psi_{\min}} \right).$$

Proof. For each τ and t, we have by Lemma E.5,

$$\| \frac{1}{\sqrt{N}} H_1' \beta' (\Omega_t - \Pi) \xi_t^\tau \| = \| \sum_{j=1}^N (\omega_{jt} - p_j) U_{M^\star, j} \xi_{t,j}^{\tau\prime} \| \leq \vartheta \max_g \| \sum_{j \in G_g} (\omega_{jt} - p_j) U_{M^\star, j} \xi_{t,j}^{\tau\prime} \|. \tag{G.13}$$

Because ξ_t^τ only depends on $M^\star$ and Y excluding the tth column of Y, conditioning on $\mathcal{M}$, $\{\omega_{jt}\}_{j \in G_g}$ are independent from ξ_t^τ. Hence, we have $\mathbb{E}[\omega_{jt} - p_j | \mathcal{M}, \xi_t^\tau] = \mathbb{E}[\omega_{jt} - p_j | \mathcal{M}] = 0$ and, conditioning on $\mathcal{M}$ and ξ_t^τ, $\{\omega_{jt} - p_j\}_{j \in G_g}$ are independent across j. Then, by matrix Bernstein inequality with the relation (G.13), we have for each τ and t that

$$\| \frac{1}{\sqrt{N}} H_1' \beta' (\Omega_t - \Pi) \xi_t^\tau \| \leq C \left(\sigma \vartheta \log T \|U_{M^\star}\|_{2,\infty} \max_t \max_\tau \|\xi_t^\tau\|_{2,\infty} + \sigma p_{\max}^{\frac{1}{2}} \vartheta \sqrt{\log T} \|U_{M^\star}\|_{2,\infty} \max_t \max_\tau \|\xi_t^\tau\|_F \right)$$

with probability exceeding $1 - O(T^{-100})$. So, we have with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$,

$$\max_t \max_\tau \| \frac{1}{\sqrt{N}} H_1' \beta' (\Omega_t - \Pi) \xi_t^\tau \| \leq C \left(\frac{\sigma p_{\max} \vartheta q^{\frac{1}{2}} \mu^{\frac{1}{2}} K \sqrt{\log T} \max\{\sqrt{N}, \sqrt{T}\}}{p_{\min} \psi_{\min}} \right).$$

for some constant $C > 0$. □

G.3 Lemmas regarding generalizations of Chen et al. (2020)

This section provides lemmas that we need for generalizing results in Chen et al. (2020) to allow the heterogeneous observation probabilities, the dependence in missing pattern, and the low-rank approximation error M^R. In stating the conditions, let us denote generic $N \times K$ and $T \times K$ matrices by $\ddot{W}$ and $\ddot{Z}$.

Condition G.1 (Regularization parameter)**.** *The regularization parameter λ satisfies (i) $\left\| \widehat{\Pi}^{-1} \mathcal{P}_\Omega(\mathcal{E}) \right\| < \frac{7}{8} \lambda$ and (ii) $\left\| \mathcal{P}_\Omega(\ddot{W}\ddot{Z}' - M^\star) - \Pi(\ddot{W}\ddot{Z}' - M^\star) \right\| < \frac{1}{88} p_{\min} \lambda$*

Condition G.2 (Low-rank approximation error)**.** *The low-rank approximation error M^R satisfies $\|\mathcal{P}_\Omega(M^R)\| \leq c\lambda$ for some sufficiently small constant $c > 0$.*

Condition G.3 (Injectivity)**.** *Let T be the tangent space of $\ddot{W}\ddot{Z}'$. There is a quantity $c_{inj} > 0$ such that $p_{\min}^{-1} \|\mathcal{P}_\Omega(H)\|_F^2 \geq c_{inj} \|H\|_F^2$ for all $H \in T$.*

Condition G.4. *$\widehat{\Pi}$, the estimator of Π, satisfies (i) $\left\| \widehat{\Pi} - \Pi \right\| \leq C \sqrt{\frac{\log N}{T}}$ for some constant $C > 0$, (ii) $\left\| \widehat{\Pi}^{-1} - \Pi^{-1} \right\| \leq C \sqrt{\frac{\log N}{T}}$ for some constant $C > 0$, (iii) $\left\| \widehat{\Pi} \right\| \leq \frac{11}{10} p_{\max}$, and (iv) $\left\| \widehat{\Pi}^{-1} \right\| \leq \frac{11}{10} p_{\min}^{-1}$.*

Lemma G.12. *Suppose that $(\ddot{W}, \ddot{Z})$ satisfies*

$$\left\| \nabla f^{\text{infs}}(\ddot{W}, \ddot{Z}) \right\|_F \leq c \frac{\sqrt{c_{\text{inj}} p_{\min}}}{q} \lambda \sqrt{\psi_{\min}} \tag{G.14}$$

for some sufficiently small constant $c > 0$, and $\left\| \ddot{W}\ddot{Z}' - M^\star \right\| \leq Cq\lambda$ for some constant $C > 0$. Additionally, assume that any nonzero singular value of $\ddot{W}$ and $\ddot{Z}$ exists in the interval $[\sqrt{\frac{\psi_{\min}}{2}}, \sqrt{2\psi_{\max}}]$, and $\sqrt{\frac{\log N}{T}} p_{\max} \ll \frac{\sqrt{c_{\text{inj}} p_{\min}}}{K^{1/2} q^{5/2}}$, $\sqrt{\frac{\log N}{T}} \ll \frac{p_{\min}}{q}$. Then, under Conditions G.1-G.4, $\widetilde{M}$, a minimizer of (2.2), satisfies $\left\| \ddot{W}\ddot{Z}' - \widetilde{M} \right\|_F \lesssim \frac{q^{5/2} K^{1/2} p_{\max}^2}{c_{\text{inj}} p_{\min}} \lambda \sqrt{\frac{\log N}{T}} + \frac{q}{p_{\min} c_{\text{inj}}} \frac{p_{\max}}{\sqrt{\psi_{\min}}} \left\| \nabla f^{\text{infs}}(\ddot{W}, \ddot{Z}) \right\|_F$.

Proof. The triangle inequality gives us

$$\left\|\nabla f^{\text{fsbl}}(\ddot{W},\ddot{Z})\right\|_F \leq \left\|\begin{bmatrix}(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)\ddot{Z}\\ [(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)]'\ddot{W}\end{bmatrix}\right\|_F + \left\|\nabla f^{\text{infs}}(\ddot{W},\ddot{Z})\right\|_F. \tag{G.15}$$

For the first term, note that

$$\left\|\begin{bmatrix}(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)\ddot{Z}\\ [(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)]'\ddot{W}\end{bmatrix}\right\|_F \leq \left\|(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)\ddot{Z}\right\|_F + \left\|[(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)]'\ddot{W}\right\|_F.$$

Also, by Conditions G.1, G.2, and the assumption $\left\|\ddot{W}\ddot{Z}'-M^\star\right\| \leq Cq\lambda$, one has $\left\|\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)\right\| \leq C_2 p_{\max} q\lambda$ for some constants $C_2>0$. Hence, Condition G.4 yields $\left\|(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)\ddot{Z}\right\|_F \leq C_3\sqrt{\frac{\log N}{T}}p_{\max}q\lambda\sqrt{K}\sqrt{\psi_{\max}}$ for some constant $C_3>0$. Similarly, we can bound $\left\|(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)\ddot{W}\right\|_F$. Combining them,

$$\left\|\begin{bmatrix}(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)\ddot{Z}\\ [(\widehat{\Pi}^{-1}-\Pi^{-1})\mathcal{P}_\Omega(\ddot{W}\ddot{Z}'-Y)]'\ddot{W}\end{bmatrix}\right\|_F \leq C\sqrt{\frac{\log N}{T}}p_{\max}q\lambda\sqrt{K}\sqrt{\psi_{\max}}$$

for some constant $C>0$. Because $\left\|\nabla f^{\text{infs}}(\ddot{W},\ddot{Z})\right\|_F \leq c_1\frac{\sqrt{c_{\text{inj}}p_{\min}}}{q}\lambda\sqrt{\psi_{\min}}$ for some sufficiently small $c_1>0$, we have $\left\|\nabla f^{\text{fsbl}}(\ddot{W},\ddot{Z})\right\|_F \leq c_2\frac{\sqrt{c_{\text{inj}}p_{\min}}}{q}\lambda\sqrt{\psi_{\min}}$ for some sufficiently small $c_2>0$. Then, by Lemma G.24, we have

$$\left\|\ddot{W}\ddot{Z}'-\widetilde{M}\right\|_F \lesssim \frac{q^{5/2}K^{1/2}p_{\max}^2}{c_{\text{inj}}p_{\min}}\lambda\sqrt{\frac{\log N}{T}} + \frac{q}{p_{\min}c_{\text{inj}}}\frac{p_{\max}}{\sqrt{\psi_{\min}}}\left\|\nabla f^{\text{infs}}(\ddot{W},\ddot{Z})\right\|_F.$$

□

Lemma G.13. *With probability at least* $1-O(\min\{N^{-101},T^{-101}\})$, *(i)* $\|\mathcal{P}_\Omega(\mathbf{1}\mathbf{1}')-\Pi\mathbf{1}\mathbf{1}'\| \lesssim \sqrt{\max\{N,T\}}$ *and (ii)* $\|\mathcal{P}_\Omega(\mathcal{E})\| \lesssim \sqrt{\max\{N,T\}}$.[1] *as long as* $C_\lambda>0$ *is sufficiently large.*

Proof. Let $Q=\mathcal{P}_\Omega(\mathbf{1}\mathbf{1}')-\Pi\mathbf{1}\mathbf{1}'$. Then Q contains independent sub-gaussian columns. Then, by Theorem 5.39 of Vershynin (2010), we have, with probability at least $1-O(\min\{N^{-101},T^{-101}\})$, $\|QQ'-\mathbb{E}[QQ']\| \lesssim \sqrt{NT}+N$. In addition, we have

$$\begin{aligned}\|\mathbb{E}[QQ']\| \leq \|\mathbb{E}[QQ']\|_1 &= \max_{1\leq j\leq N}\sum_{i=1}^N\left|(\mathbb{E}[QQ'])_{ij}\right| = \max_{1\leq j\leq N}\sum_{i=1}^N\left|\sum_{t=1}^T\text{Cov}(\omega_{it},\omega_{jt})\right|\\ &\leq \max_{1\leq j\leq N}\sum_{i=1}^N\sum_{t=1}^T|\text{Cov}(\omega_{it},\omega_{jt})| \leq T\max_{1\leq t\leq T}\max_{1\leq j\leq N}\sum_{i=1}^N|\text{Cov}(\omega_{it},\omega_{jt})| \lesssim T,\end{aligned}$$

where $\|\cdot\|_1$ means Schatten 1-norm. The last relation uses Assumption 3.3 (i). Therefore, with probability at least $1-O(\min\{N^{-101},T^{-101}\})$, we have $\|QQ'\| \lesssim \max\{N,T\}$, and thus $\|Q\| \lesssim \sqrt{\max\{N,T\}}$. Similarly, we can show $\|\mathcal{P}_\Omega(\mathcal{E})\| \lesssim \sqrt{\max\{N,T\}}$ by setting $Q=\mathcal{P}_\Omega(\mathcal{E})$. □

Lemma G.14. *Assume that* $\max_{it}|M_{it}^R| \ll 1/\max\{\sqrt{N},\sqrt{T}\}$. *Then* $\left\|\mathcal{P}_\Omega(M^R)\right\| \leq c\min\{\sqrt{N},\sqrt{T}\}$ *for some sufficiently small* $c>0$. *It implies that Condition G.2 holds if* $\lambda = C_\lambda\sigma\sqrt{\max\{N,T\}\frac{p_{\max}}{p_{\min}^2}}$ *for some constant* $C_\lambda>0$.

Proof. It comes from the following relation $\left\|\mathcal{P}_\Omega(M^R)\right\| \leq \left\|\mathcal{P}_\Omega(M^R)\right\|_F \leq \left\|M^R\right\|_F = \sqrt{\sum_{i,t}(M_{it}^R)^2} \leq \sqrt{NT}\max_{it}|M_{it}^R| \leq c\min\{\sqrt{N},\sqrt{T}\}$ for some sufficiently small $c>0$. □

Lemma G.15. *The followings show that Condition G.4 hold with high probability.*

[1] This bound and Lemma G.15 (iv) together imply that Condition G.1 (i) holds if $\lambda = C_\lambda\sigma\sqrt{\max\{N,T\}\frac{p_{\max}}{p_{\min}^2}}$

(i) There is a constant $C > 0$ such that $\left\|\widehat{\Pi} - \Pi\right\| \le C\sqrt{\frac{\log N}{T}}$ with probability $1 - 2N^{-7}$.

(ii) In the event of (i), if $C\sqrt{\frac{\log N}{T}} \le \frac{1}{2}p_{\min}$, then there is a constant $C_1 > 0$ such that $\left\|\widehat{\Pi}^{-1} - \Pi^{-1}\right\| \le C_1\sqrt{\frac{\log N}{T}}$.

(iii) In the event of (i), if $C\sqrt{\frac{\log N}{T}} \le \frac{1}{10}p_{\max}$, then $\left\|\widehat{\Pi}\right\| \le \frac{11}{10}p_{\max}$.

(iv) In the event of (i), if $C_1\sqrt{\frac{\log N}{T}} \le \frac{1}{10}p_{\min}^{-1}$, then $\left\|\widehat{\Pi}^{-1}\right\| \le \frac{11}{10}p_{\min}^{-1}$.

Proof. For the first result, we use the Hoeffding's inequality for bounded random variables. (ex., Theorem 2.2.6 of Vershynin (2010)) Note that $\{\omega_{it}\}$ are independent Bernoulli random variables with the expectation p_i. Then, by Hoeffding's inequality for any $s > 0$, $P\left(\sqrt{T}|\widehat{p}_i - p_i| \ge s\right) = P\left(\left|\sum_{t=1}^T \frac{1}{\sqrt{T}}(\omega_{it} - p_i)\right| \ge s\right) \le 2\exp(-2s^2)$ for all i. Then, we have

$$P\left(\max_{1\le i\le N} |\widehat{p}_i - p_i| \ge \frac{s}{\sqrt{T}}\right) = P\left(\max_{1\le i\le N} \sqrt{T}|\widehat{p}_i - p_i| \ge s\right) = P\left(\bigcup_{1\le i\le N}\{\sqrt{T}|\widehat{p}_i - p_i| \ge s\}\right)$$
$$\le \sum_{1\le i\le N} P\left(\sqrt{T}|\widehat{p}_i - p_i| \ge s\right) \le 2N\exp(-2s^2).$$

By taking $s = 2\sqrt{\log N}$, we have that $P\left(\max_{1\le i\le N}|\widehat{p}_i - p_i| \ge \frac{2\sqrt{\log N}}{\sqrt{T}}\right) \le \frac{2}{N^7}$. Therefore, we have $P\left(\left\|\widehat{\Pi} - \Pi\right\| \le \frac{2\sqrt{\log N}}{\sqrt{T}}\right) = P\left(\max_{1\le i\le N}|\widehat{p}_i - p_i| \le \frac{2\sqrt{\log N}}{\sqrt{T}}\right) \le 1 - \frac{2}{N^7}$. For (ii), (iii), and (iv), the proofs follow from straightforward basic inequalities, so are omitted.

□

G.4 Extension of Lemmas in Chen et al. (2020)

This section includes lemmas from Chen et al. (2020) which are used in the proof of Lemma E.2. Although our setting is more general than Chen et al. (2020), some of the following lemmas have fairly similar proofs as their counterparts in Chen et al. (2020). For the sake of brevity, we will omit some proofs. The omitted proofs can be provided upon request. Here, we include the proofs of Lemma G.17 and Lemma G.18 since their proofs handle the clustered dependence in missing pattern, which incurs nontrivial modifications from proofs in Chen et al. (2020).

We first establish Lemma G.16 in an induction manner. We plan to show that

$$\|\mathcal{F}^\tau H^\tau - \mathcal{F}\|_F \le C_F\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|W\|_F, \tag{G.16}$$

$$\|\mathcal{F}^\tau H^\tau - \mathcal{F}\| \le C_{op}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|W\|, \tag{G.17}$$

$$\max_{1\le l\le N+T}\left\|\mathcal{F}^\tau H^\tau - \mathcal{F}^{\tau,(l)}Q^{\tau,(l)}\right\|_F \le C_3\sqrt{\vartheta}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|\mathcal{F}\|_{2,\infty}, \tag{G.18}$$

$$\max_{1\le l\le N+T}\left\|\left(\mathcal{F}^{\tau,(l)}H^{\tau,(l)} - \mathcal{F}\right)_{l,\cdot}\right\|_2 \le C_4\sqrt{\vartheta}q\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|\mathcal{F}\|_{2,\infty}, \tag{G.19}$$

$$\|\mathcal{F}^\tau H^\tau - \mathcal{F}\|_{2,\infty} \le C_\infty\sqrt{\vartheta}q\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|\mathcal{F}\|_{2,\infty}, \tag{G.20}$$

$$\|W^{\tau\prime}W^\tau - Z^{\tau\prime}Z^\tau\|_F \le C_B q\eta\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\sqrt{K}\psi_{\max}^2 \tag{G.21}$$

hold for all $0 \leq \tau \leq \tau_0 = \max\{N^{18}, T^{18}\}$ and for some constants C_F, C_{op}, C_3, C_4, C_∞, $C_B > 0$, provided that $\eta \overset{c}{\asymp} \frac{1}{\max\{N,T\}q^3\psi_{\max}}$. We will additionally show that

$$f^{\text{infs},(l)}(W^{\tau,(l)}, Z^{\tau,(l)}) \leq f^{\text{infs},(l)}(W^{\tau-1,(l)}, Z^{\tau-1,(l)}) - \frac{\eta}{2} \left\| \nabla f^{\text{infs},(l)}(W^{\tau-1,(l)}, Z^{\tau-1,(l)}) \right\|_F^2 \tag{G.22}$$

hold for all $1 \leq l \leq N+T$ and for all $1 \leq \tau \leq \tau_0 = \max\{N^{18}, T^{18}\}$.

Lemma G.16. *With probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$, the iterates $\{(W^\tau, Z^\tau)\}_{0\leq\tau\leq\tau_0}$ of (D.3) and $\{(W^{\tau,(l)}, Z^{\tau,(l)})\}_{0\leq\tau\leq\tau_0}$, for all $1 \leq l \leq N+T$, of (D.4) satisfy the following:*

$$\max\left\{ \|W^\tau H^\tau - W\|_F, \|Z^\tau H^\tau - Z\|_F \right\} \leq C_F \left(\frac{\sigma}{\psi_{\min}} \sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}} \right) \|W\|_F, \tag{G.23}$$

$$\max\left\{ \|W^\tau H^\tau - W\|, \|Z^\tau H^\tau - Z\| \right\} \leq C_{op} \left(\frac{\sigma}{\psi_{\min}} \sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}} \right) \|W\|, \tag{G.24}$$

$$\begin{aligned}&\max\left\{ \|W^\tau H^\tau - W\|_{2,\infty}, \|Z^\tau H^\tau - Z\|_{2,\infty} \right\} \\ &\quad\leq C_\infty \sqrt{\vartheta} q \left(\frac{\sigma}{\psi_{\min}} \sqrt{\frac{\max\{N\log N, T\log T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}} \right) \max\{\|W\|_{2,\infty}, \|Z\|_{2,\infty}\},\end{aligned} \tag{G.25}$$

$$\begin{aligned}&\max_{1\leq l\leq N+T} \max\left\{ \left\|W^\tau H^\tau - W^{\tau,(l)} H^{\tau,(l)}\right\|_F, \left\|Z^\tau H^\tau - Z^{\tau,(l)} H^{\tau,(l)}\right\|_F \right\} \\ &\quad\leq C_{l_0} \sqrt{\vartheta} q \left(\frac{\sigma}{\psi_{\min}} \sqrt{\frac{\max\{N\log N, T\log T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}} \right) \max\{\|W\|_{2,\infty}, \|Z\|_{2,\infty}\},\end{aligned} \tag{G.26}$$

$$\begin{aligned}&\max_{1\leq l\leq N+T} \max\left\{ \left\|W^{\tau,(l)} H^{\tau,(l)} - W\right\|_{2,\infty}, \left\|Z^{\tau,(l)} H^{\tau,(l)} - Z\right\|_{2,\infty} \right\} \\ &\quad\leq 5(C_3 + C_\infty) \sqrt{\vartheta} q \left(\frac{\sigma}{\psi_{\min}} \sqrt{\frac{\max\{N\log N, T\log T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}} \right) \max\{\|W\|_{2,\infty}, \|Z\|_{2,\infty}\},\end{aligned} \tag{G.27}$$

$$\|W^\tau Z^{\tau\prime} - M^\star\|_F \leq 3qC_F \left(\frac{\sigma}{\psi_{\min}} \sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}} \right) \|M^\star\|_F, \tag{G.28}$$

$$\|W^\tau Z^{\tau\prime} - M^\star\| \leq 3C_{op} \left(\frac{\sigma}{\psi_{\min}} \sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}} \right) \|M^\star\|, \tag{G.29}$$

where $C_F, C_{op}, C_\infty, C_{l_0} > 0$ are some absolute constants which do not depend on τ.

Proof. First, note that (G.16) - (G.21) hold when $\tau = 0$ because $\mathcal{F}^0 = \mathcal{F}^{0,(l)} = \mathcal{F}$ for all $1 \leq l \leq N+T$. Then, from the mathematical induction using Lemma G.17, G.18, G.19 and Lemma G.20 , with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$, (G.16) - (G.21) hold for all $0 \leq \tau \leq \tau_0$. Then, (G.23), (G.24) and (G.25) are immediate results of it. In addition, by Lemma G.22 (iii) with the fact that (G.18) holds for all $0 \leq \tau \leq \tau_0$ with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$, we can have (G.26). For (G.27), note that

$$\max_{1\leq l\leq N+T} \max\left\{ \left\|W^{\tau,(l)} H^{\tau,(l)} - W\right\|_{2,\infty}, \left\|Z^{\tau,(l)} H^{\tau,(l)} - Z\right\|_{2,\infty} \right\} \leq \max_{1\leq l\leq N+T} 5q \left\|\mathcal{F}^{\tau,(l)} Q^{\tau,(l)} - \mathcal{F}^\tau H^\tau\right\|_F + \|\mathcal{F}^\tau H^\tau - \mathcal{F}\|_{2,\infty}$$

by Lemma G.22. Because (G.18) and (G.20) hold for all $0 \leq \tau \leq \tau_0$ with probability at least $1 - O(\min\{N^{-5}, T^{-5}\})$, we have for all $0 \leq \tau \leq \tau_0$,

$$\max_{1\leq l\leq N+T} \max\left\{ \left\|W^{\tau,(l)} H^{\tau,(l)} - W\right\|_{2,\infty}, \left\|Z^{\tau,(l)} H^{\tau,(l)} - Z\right\|_{2,\infty} \right\}$$

$$\le 5(C_3+C_\infty)\sqrt{\vartheta}q\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N,T\log T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\max\{\|W\|_{2,\infty},\|Z\|_{2,\infty}\}.$$

with probability at least $1-O(\min\{N^{-5},T^{-5}\})$. For (G.28), note the following decomposition $W^\tau Z^{\tau\prime}-M^\star=(W^\tau H^\tau-W)(Z^\tau H^\tau)'+W(Z^\tau H^\tau-Z)'$, which together with the triangle inequality gives $\|W^\tau Z^{\tau\prime}-M^\star\|_F\le\|W^\tau H^\tau-W\|_F\|Z^\tau H^\tau\|+\|W\|\|Z^\tau H^\tau-Z\|_F$. Considering Lemma G.22, we have $\|Z^\tau H^\tau\|\le 2\|W\|$. Because (G.16) holds for all $0\le\tau\le\tau_0$ with probability at least $1-O(\min\{N^{-5},T^{-5}\})$, we have for all $0\le\tau\le\tau_0$, $\|W^\tau Z^{\tau\prime}-M^\star\|_F\le 3qC_F\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\|M^\star\|_F$ with probability at least $1-O(\min\{N^{-5},T^{-5}\})$. We can derive (G.29) similarly. □

Lemma G.17. *Suppose that $\lambda=C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ for some large constant $C_\lambda>0$, $0<\eta\ll 1/(q^2\psi_{\max}\min\{N,T\})$. Suppose also that the iterates satisfy (G.16)-(G.21) at the τth iteration, then with probability at least $1-O(\min\{N^{-99},T^{-99}\})$, we have*

$$\max_{1\le l\le N+T}\left\|\mathcal{F}^{\tau+1}H^{\tau+1}-\mathcal{F}^{\tau+1,(l)}Q^{\tau+1,(l)}\right\|_F\le C_3\sqrt{\vartheta}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N,T\log T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\|\mathcal{F}\|_{2,\infty}$$

where C_3 is some sufficiently large constant.

Lemma G.18. *Suppose that $\lambda=C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ for some large constant $C_\lambda>0$, and $0<\eta\ll 1/(q^2\sqrt{K}\psi_{\max})$. Suppose also that the iterates satisfy (G.16)-(G.21) at the τth iteration, then with probability at least $1-O(\min\{N^{-99},T^{-99}\})$,*

$$\max_{1\le l\le N+T}\left\|\left(\mathcal{F}^{\tau+1,(l)}H^{\tau+1,(l)}-\mathcal{F}\right)_{l,\cdot}\right\|_2\le C_4\sqrt{\vartheta}q\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N,T\log T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\|\mathcal{F}\|_{2,\infty}.$$

Lemma G.19. *Suppose that $\lambda=C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ for some large constant $C_\lambda>0$, and $0<\eta\ll\min\{1/(q^{5/2}\psi_{\max}),1/(q^3\psi_{\max}\sqrt{K})$ Suppose also that the iterates satisfy (G.16)-(G.21) at the τth iteration, then with probability at least $1-O(\min\{N^{-100},T^{-100}\})$,*

$$\left\|\mathcal{F}^{\tau+1}H^{\tau+1}-\mathcal{F}\right\|_F\le C_F\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\|W\|_F,$$

$$\left\|\mathcal{F}^{\tau+1}H^{\tau+1}-\mathcal{F}\right\|\le C_{op}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\|W\|,$$

$$\left\|W^{\tau+1\prime}W^{\tau+1}-Z^{\tau+1\prime}Z^{\tau+1}\right\|_F\le C_Bq\eta\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\sqrt{K}\psi_{\max}^2,$$

$$\max_{1\le l\le N+T}\left\|W^{\tau+1,(l)\prime}W^{\tau+1,(l)}-Z^{\tau+1,(l)\prime}Z^{\tau+1,(l)}\right\|_F\le C_Bq\eta\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\sqrt{K}\psi_{\max}^2,$$

where $C_F>0$ and C_{op} are sufficiently large, and $C_B\gg C_{op}^2$.

Lemma G.20. *Suppose that $\lambda=C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ for some large constant $C_\lambda>0$. Suppose also that the iterates satisfy (G.16)-(G.21) at the τth iteration, then with probability at least $1-O(\min\{N^{-98},T^{-98}\})$,*

$$\left\|\mathcal{F}^{\tau+1}H^{\tau+1}-\mathcal{F}\right\|_{2,\infty}\le C_\infty\sqrt{\vartheta}q\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N,T\log T\}p_{\max}}{p_{\min}^2}}+\frac{\lambda}{\psi_{\min}}\right)\|\mathcal{F}\|_{2,\infty}$$

holds as long as $C_\infty\ge 5C_3+C_4$.

Lemma G.21. *Suppose that $\lambda=C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ for some large constant $C_\lambda>0$, and $0<\eta\ll 1/(q\psi_{\max}\max\{N,T\})$. Suppose also that the iterates satisfy (G.16)-(G.21) at the τth iteration, then with probability at least $1-O(\min\{N^{-98},T^{-98}\})$,* *for all* $1 \le l \le N+T$,

$$f^{\text{infs},(l)}(W^{\tau+1,(l)}, Z^{\tau+1,(l)}) \le f^{\text{infs},(l)}(W^{\tau,(l)}, Z^{\tau,(l)}) - \frac{\eta}{2}\left\|\nabla f^{\text{infs},(l)}(W^{\tau,(l)}, Z^{\tau,(l)})\right\|_F^2.$$

Lemma G.22. *Throughout the set of results, we assume that the* τ*th iterates satisfy the induction hypotheses* (G.16)-(G.21).

(i) Suppose that $\min\{N,T\} \gg \vartheta\mu K \max\{\log N, \log T\}$. *Then, we obtain*

$$\left\|\mathcal{F}^{\tau,(l)}Q^{\tau,(l)} - \mathcal{F}\right\|_{2,\infty} \le \sqrt{\vartheta}(C_\infty q + C_3)\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N, T\log T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|\mathcal{F}\|_{2,\infty}, \tag{G.30}$$

$$\left\|\mathcal{F}^{\tau,(l)}Q^{\tau,(l)} - \mathcal{F}\right\| \le 2C_{op}\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} + \frac{\lambda}{\psi_{\min}}\right)\|W\|. \tag{G.31}$$

(ii) Suppose that $\sqrt{\vartheta}\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} \ll \frac{1}{q\sqrt{\max\{\log N, \log T\}}}$. *Then, we have*

$$\|\mathcal{F}^\tau H^\tau - \mathcal{F}\| \le \|W\|, \quad \|\mathcal{F}^\tau H^\tau - \mathcal{F}\|_F \le \|W\|_F, \quad \|\mathcal{F}^\tau H^\tau - \mathcal{F}\|_{2,\infty} \le \|\mathcal{F}\|_{2,\infty}, \tag{G.32}$$

$$\|\mathcal{F}^\tau\| \le 2\|W\|, \quad \|\mathcal{F}^\tau\|_F \le 2\|W\|_F, \quad \|\mathcal{F}^\tau\|_{2,\infty} \le 2\|\mathcal{F}\|_{2,\infty}. \tag{G.33}$$

(iii) Suppose that $\sqrt{\vartheta}\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} \ll \frac{1}{q\sqrt{\max\{\log N, \log T\}}}$ *and* $\sqrt{\frac{\mu K}{\min\{N,T\}}} \ll 1$. *Then, we have* $\left\|\mathcal{F}^\tau H^\tau - \mathcal{F}^{\tau,(l)}H^{\tau,(l)}\right\|_F \le 5q\left\|\mathcal{F}^\tau H^\tau - \mathcal{F}^{\tau,(l)}Q^{\tau,(l)}\right\|_F$.

(iv) Suppose that $\sqrt{\vartheta}\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}} \ll \frac{1}{q\sqrt{\max\{\log N, \log T\}}}$ *and* $\min\{N,T\} \ge q\mu$. *Then* (G.32), (G.33) *also hold for* $\mathcal{F}^{\tau,(l)}H^{\tau,(l)}$. *Additionally, we have* $[\psi_{\min}/2 \le \psi_{\min}\left((Z^{\tau,(l)}H^{\tau,(l)})'Z^{\tau,(l)}H^{\tau,(l)}\right) \le \psi_{\max}\left((Z^{\tau,(l)}H^{\tau,(l)})'Z^{\tau,(l)}H^{\tau,(l)}\right) \le 2\psi_{\max}$.

Lemma G.23. *Suppose that* $\eta \overset{c}{\asymp} 1/\max\{N,T\}q^3\psi_{\max}$. *Then we have, with probability at least* $1 - O(\min\{N^{-5}, T^{-5}\})$, *there is a constant* $C_{gr} > 0$ *such that* $\min_{0\le\tau<\tau_0}\left\|\nabla f^{\text{infs}}(W^\tau, Z^\tau)\right\|_F \le C_{gr}\frac{1}{\max\{N^5,T^5\}}\lambda\sqrt{\psi_{\min}}$

The following two lemmas are used in Section G.3.

Lemma G.24. *Suppose that* $(\ddot{W}, \ddot{Z})$ *satisfies* $\left\|\nabla f^{\text{fsbl}}(\ddot{W}, \ddot{Z})\right\|_F \le c\frac{\sqrt{c_{\text{inj}}p_{\min}}}{q}\lambda\sqrt{\psi_{\min}}$ *for some sufficiently small constant* $c > 0$, *and* $\left\|\ddot{W}\ddot{Z}' - M^\star\right\| \le Cq\lambda$ *for some constant* $C > 0$. *Let* $\sqrt{\frac{\log N}{T}} \ll \frac{p_{\min}}{q}$. *Additionally, assume that any nonzero singular value of* $\ddot{W}$ *and* $\ddot{Z}$ *exists in the interval* $[\sqrt{\frac{\psi_{\min}}{2}}, \sqrt{2\psi_{\max}}]$. *Then, under Conditions G.1-G.4,* $\widetilde{M}$, *a minimizer of* (2.2), *satisfies* $\left\|\ddot{W}\ddot{Z}' - \widetilde{M}\right\|_F \lesssim \frac{q}{p_{\min}c_{inj}}\frac{p_{\max}}{\sqrt{\psi_{\min}}}\left\|\nabla f^{\text{fsbl}}(\ddot{W}, \ddot{Z})\right\|_F$.

Lemma G.25. *Assume* $\sqrt{\vartheta}\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N^2,T^2\}p_{\max}}{\min\{N,T\}p_{\min}^2}} \ll \frac{1}{\sqrt{q^4\mu K\max\{\log N, \log T\}}}$. *Also, assume that* $p_{\min}^2\min\{N^2,T^2\} \gg \vartheta^2 q^4\mu^2K^2p_{\max}\max\{N\log N, T\log T\}$. *Assume that* $\lambda = C_\lambda\sigma\sqrt{\frac{\max\{N,T\}p_{\max}}{p_{\min}^2}}$ *for some large constant* $C_\lambda > 0$. *Further, let* T *denote the tangent space of* $\ddot{W}\ddot{Z}'$. *Then, with probability at least* $1 - O(\min\{N^{-100}, T^{-100}\}$,

$$\left\|\mathcal{P}_\Omega(\ddot{W}\ddot{Z}' - M^\star) - \Pi(\ddot{W}\ddot{Z}' - M^\star)\right\| < \frac{1}{88}\lambda p_{\min} \quad \textit{(Condition G.1 (ii))}$$

$$p_{\min}^{-1}\|\mathcal{P}_\Omega(H)\|_F^2 \ge \left\|\Pi^{-1/2}\mathcal{P}_\Omega(H)\right\|_F^2 \ge \frac{1}{32q}\|H\|_F^2 \quad \textit{for all } H \in T \quad \textit{(Condition G.3 with } c_{\text{inj}} = 1/(32q)\textit{)}$$

hold uniformly for all $(\ddot{W}, \ddot{Z})$ *satisfying*

$$\max\left\{\left\|\ddot{W}-W\right\|_{2,\infty},\left\|\ddot{Z}-Z\right\|_{2,\infty}\right\}\leq C\sqrt{\vartheta}q\left(\frac{\sigma}{\psi_{\min}}\sqrt{\frac{\max\{N\log N,T\log T\}p_{\max}}{p_{\min}^{2}}}+\frac{\lambda}{\psi_{\min}}\right)\max\left\{\|W\|_{2,\infty},\|Z\|_{2,\infty}\right\} \tag{G.34}$$

for some constant $C>0$.

Proof. Lemma G.17 - G.25 are simple extensions of Chen et al. (2020). So, we omit the proof for simplicity. □

References

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

Chen, Y., Y. Chi, J. Fan, C. Ma, and Y. Yan (2020): "Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization," *SIAM journal on optimization*, 30, 3098–3121.

Chernozhukov, V., C. Hansen, Y. Liao, and Y. Zhu (2021): "Inference for low-rank models," *arXiv preprint arXiv:2107.02602*.

Koltchinskii, V., K. Lounici, A. B. Tsybakov, et al. (2011): "Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion," *The Annals of Statistics*, 39, 2302–2329.

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

Yu, Y., T. Wang, and R. J. Samworth (2015): "A useful variant of the Davis–Kahan theorem for statisticians," *Biometrika*, 102, 315–323.